Applied and Environmental Microbiology

Supplementary Information

Ascomycota *Aspergillus oryzae* is an efficient expression host for production of Basidiomycota terpenes using genomic DNA sequences

Shota Nagamine, Chengwei Liu, Jumpei Nishishita, Takuto Kozaki, Kaho Sogahata, Yoshiro Sato, Atsushi Minami*, Taro Ozaki, Claudia Schmidt-Dannert, Junichi Maruyama and Hideaki Oikawa*

(A)

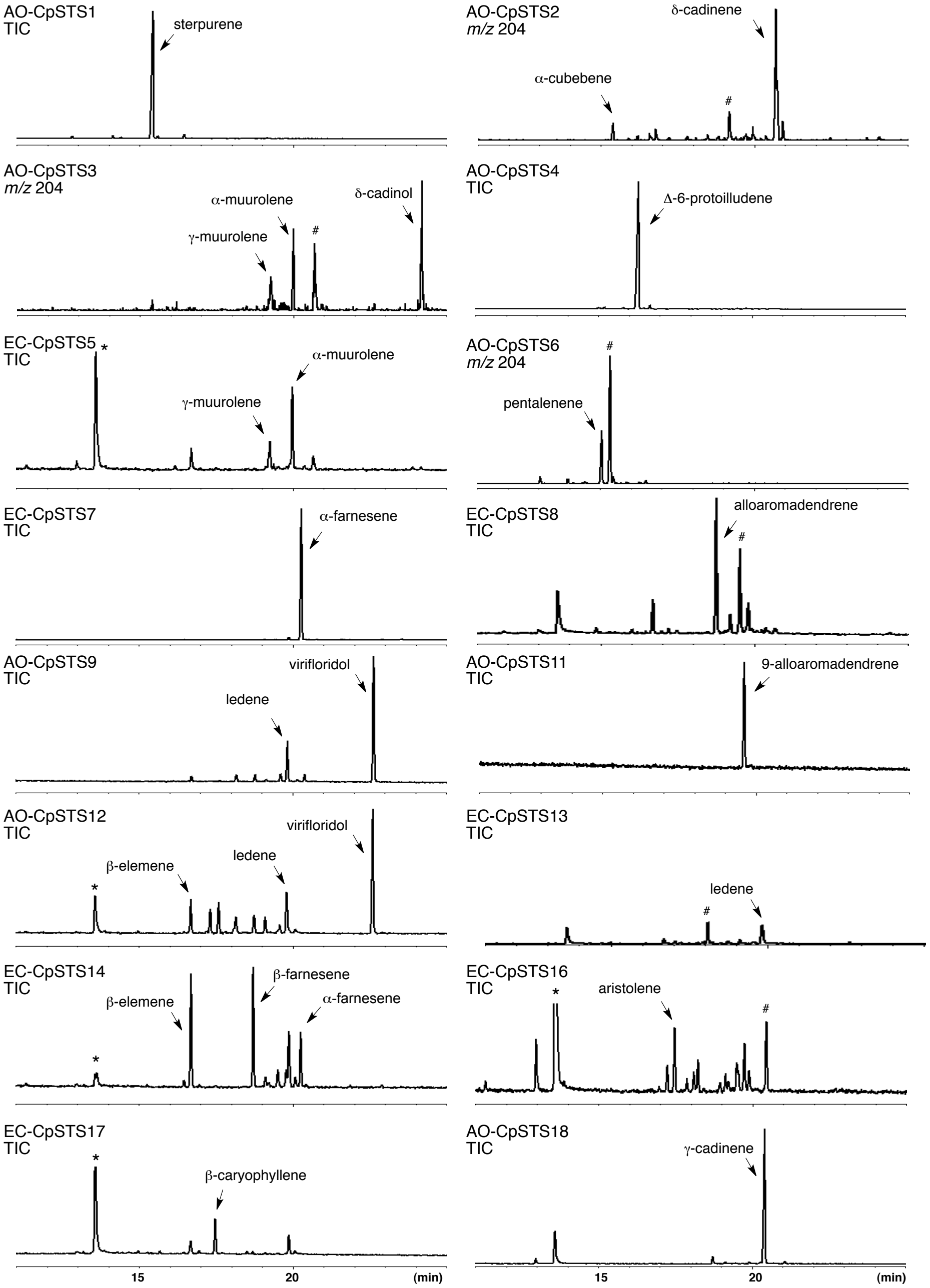


Figure S1. GC-MS profiles of sesquiterpenes produced by (A) CpSTSs and (B) ShSTSs. *Aspergillus oryzae* (AO) transformants harbor the genomic DNA sequence and *Escherichia coli* (EC) transformants do the corrected cDNA sequence. # shows an uncharacterized sesquiterpene. * shows indole produced by *E. coli* cells. Mass spectra of the sesquiterpenes obtained from (C) transformants harboring CpSTS genes and (D) those harboring ShSTS genes. Retention time is shown in the parenthesis.

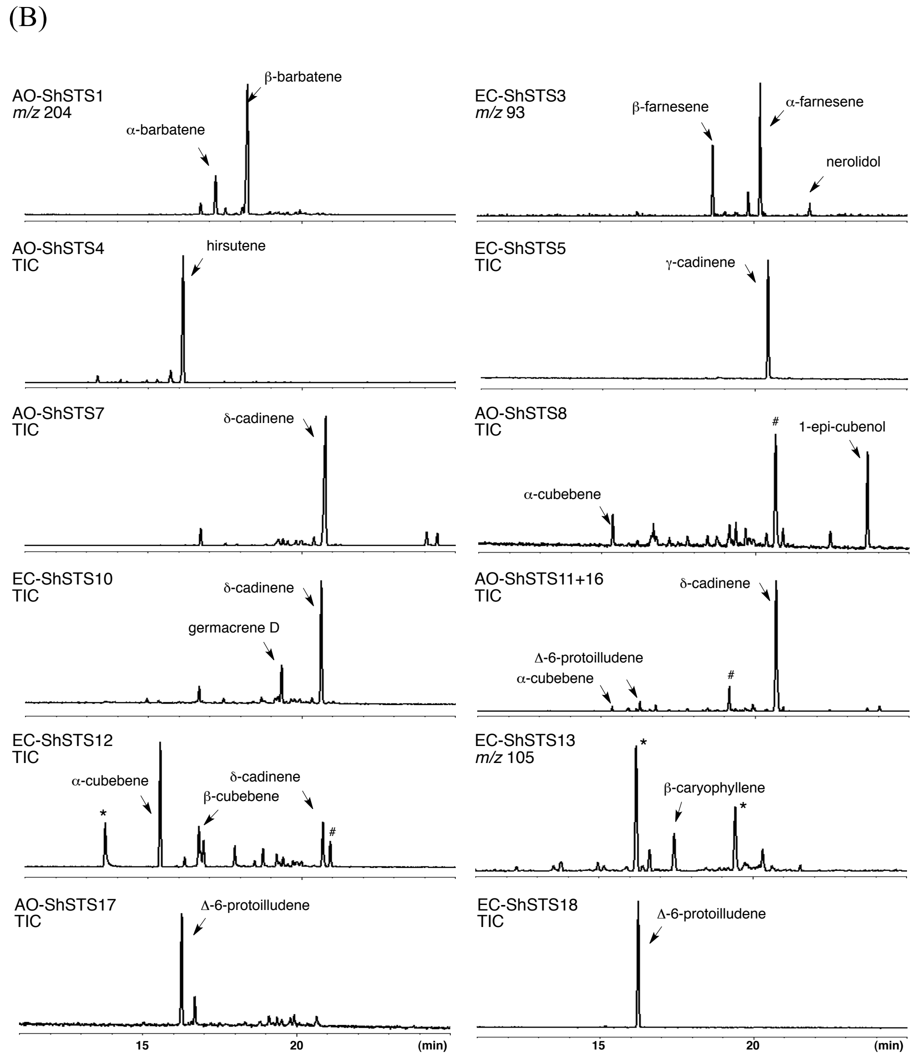


Figure S1. Continued.

(C)

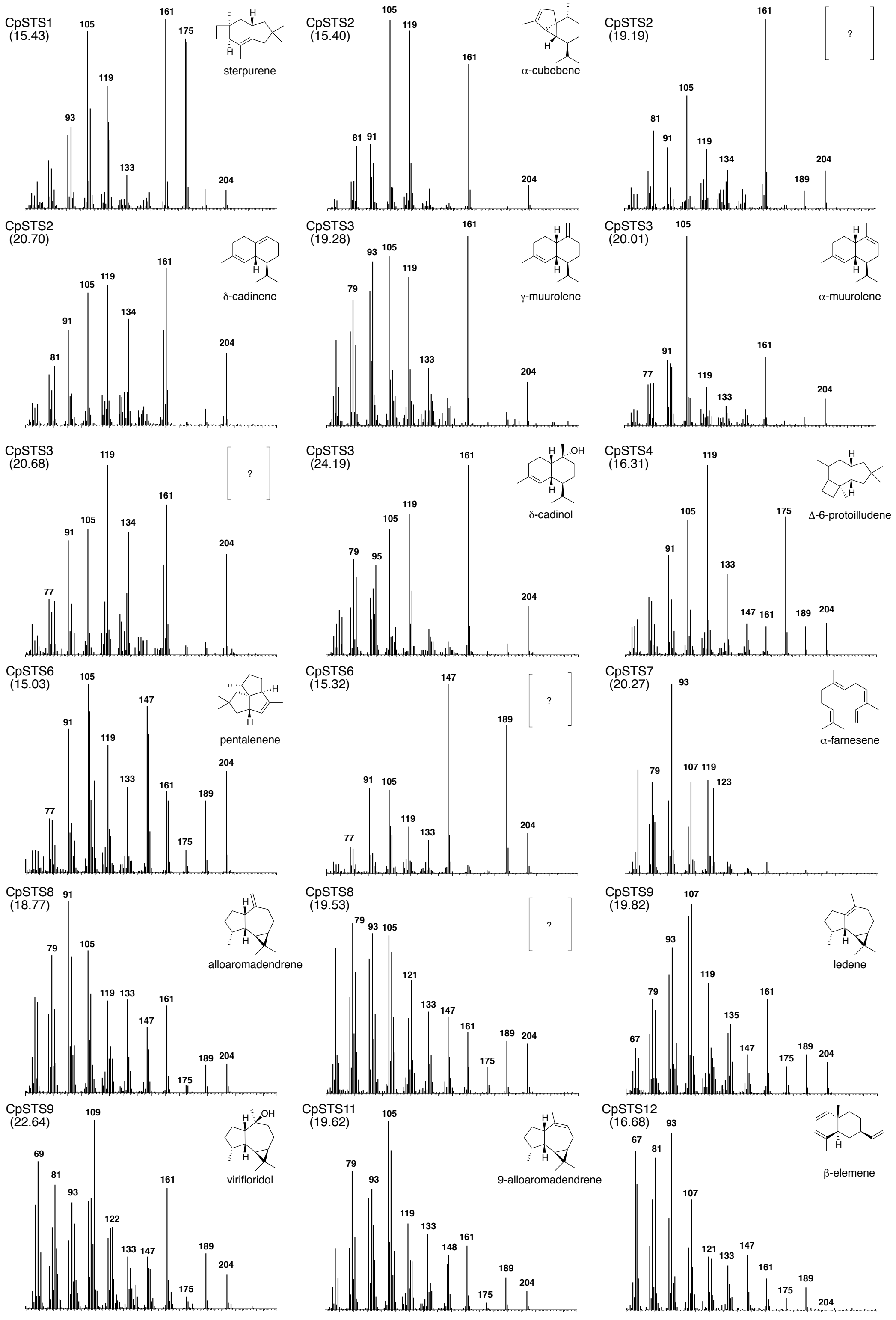


Figure S1. Continued.

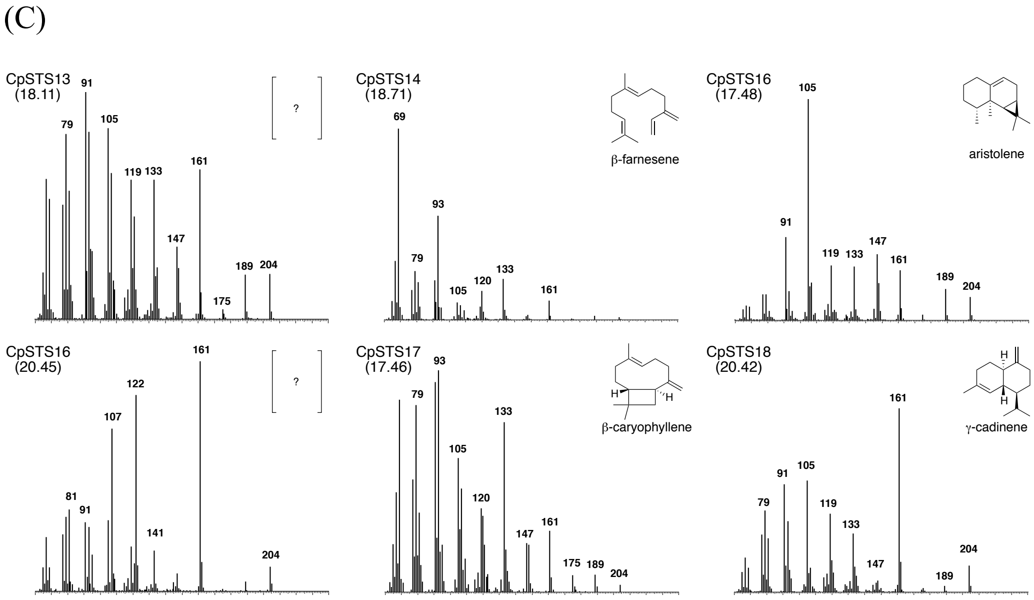


Figure S1. Continued.

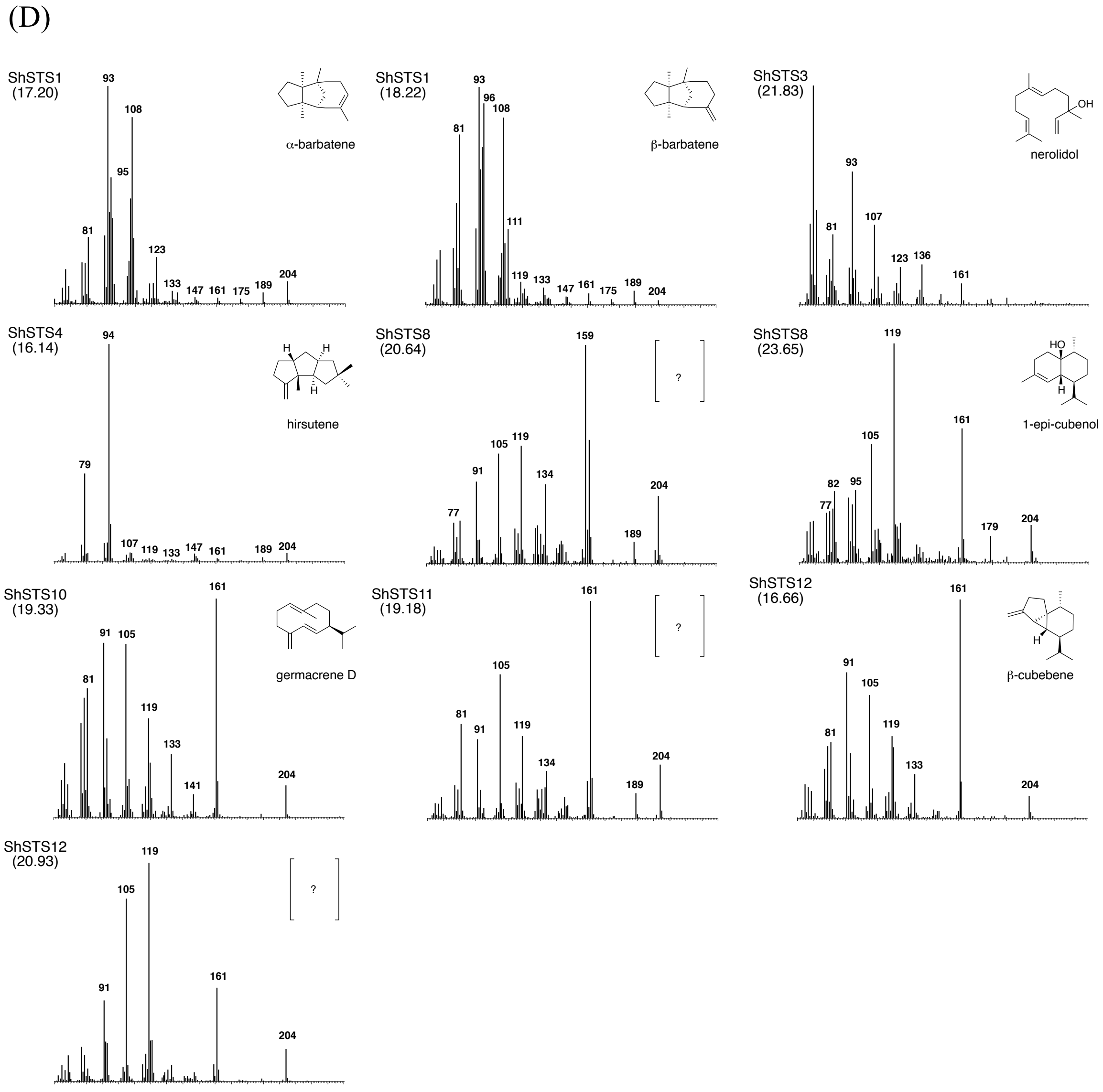


Figure S1. Continued.

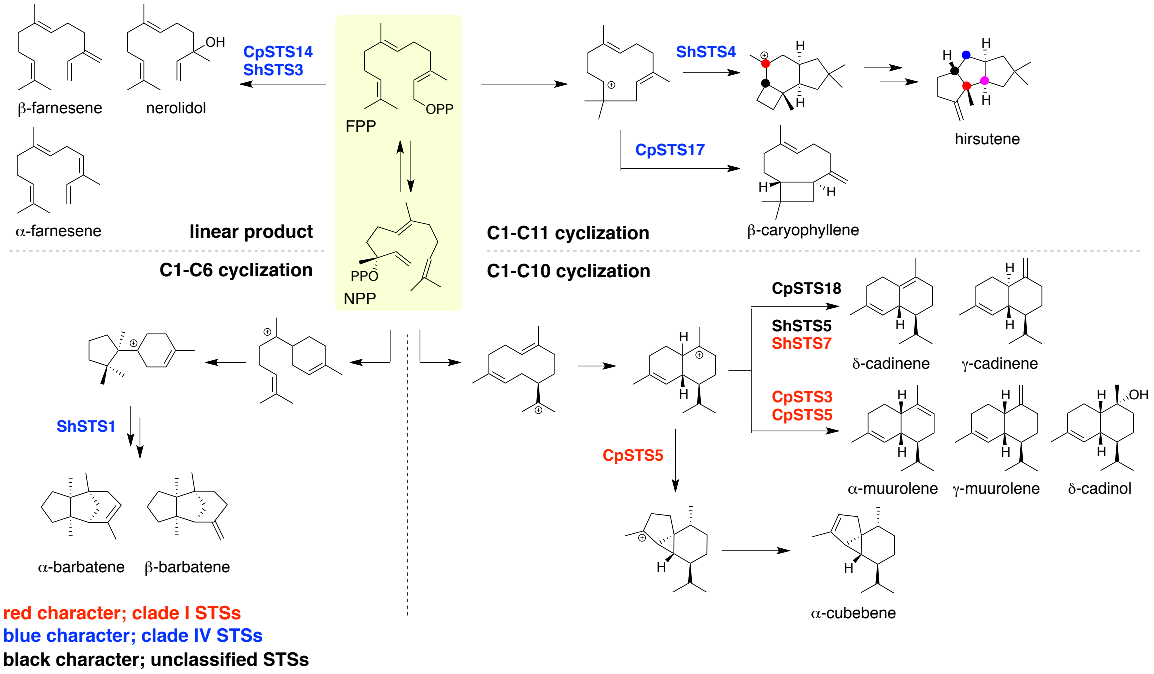


Figure S2. Proposed cyclization mechanisms of sesquiterpenes produced by clades I and IV STSs and unclassified STSs.

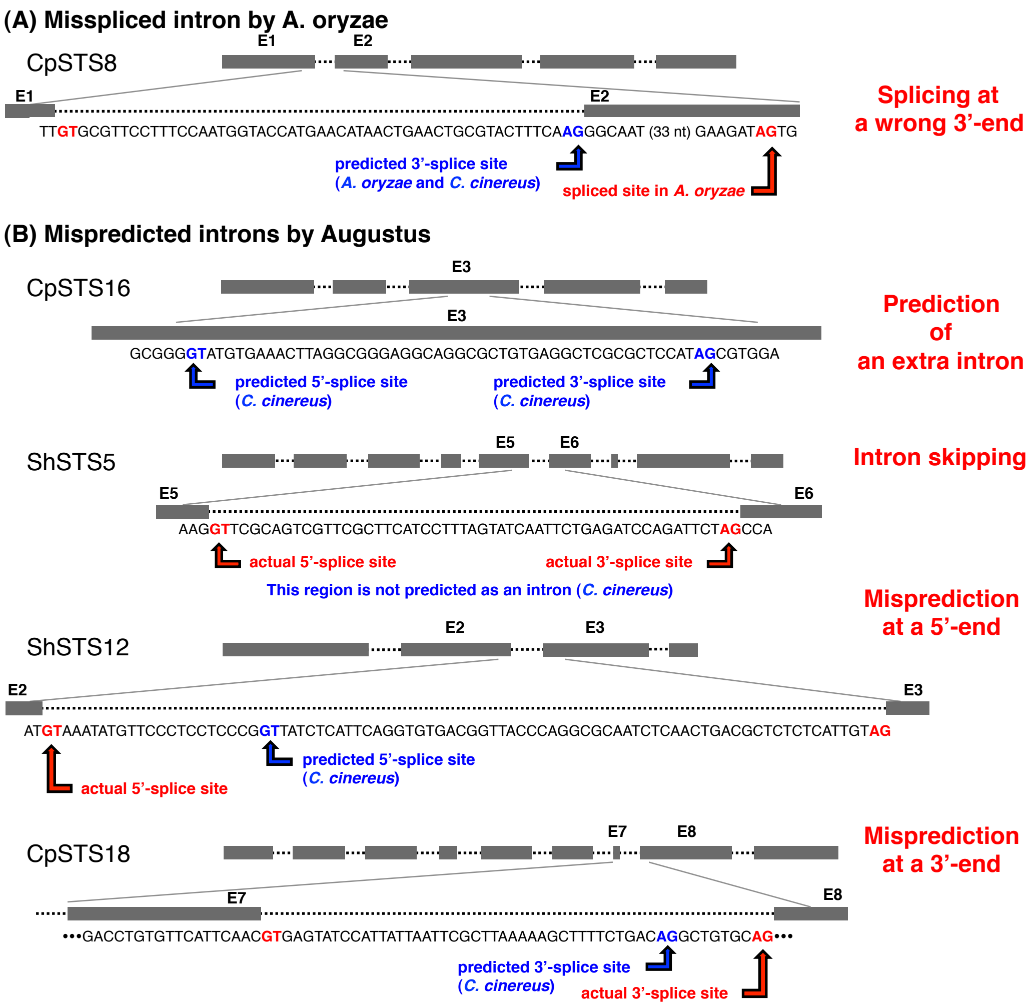


Figure S3. Schematic view of (A) misspliced intron (CpSTS8-intorn 1) in *A. oryzae* and (B) mispredicted introns by Augustus (*C. cinereus* prediction).

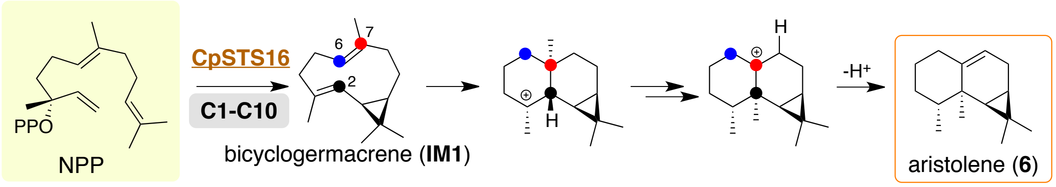


Fig S4. Alternative cyclization mechanism leading to aristolene.

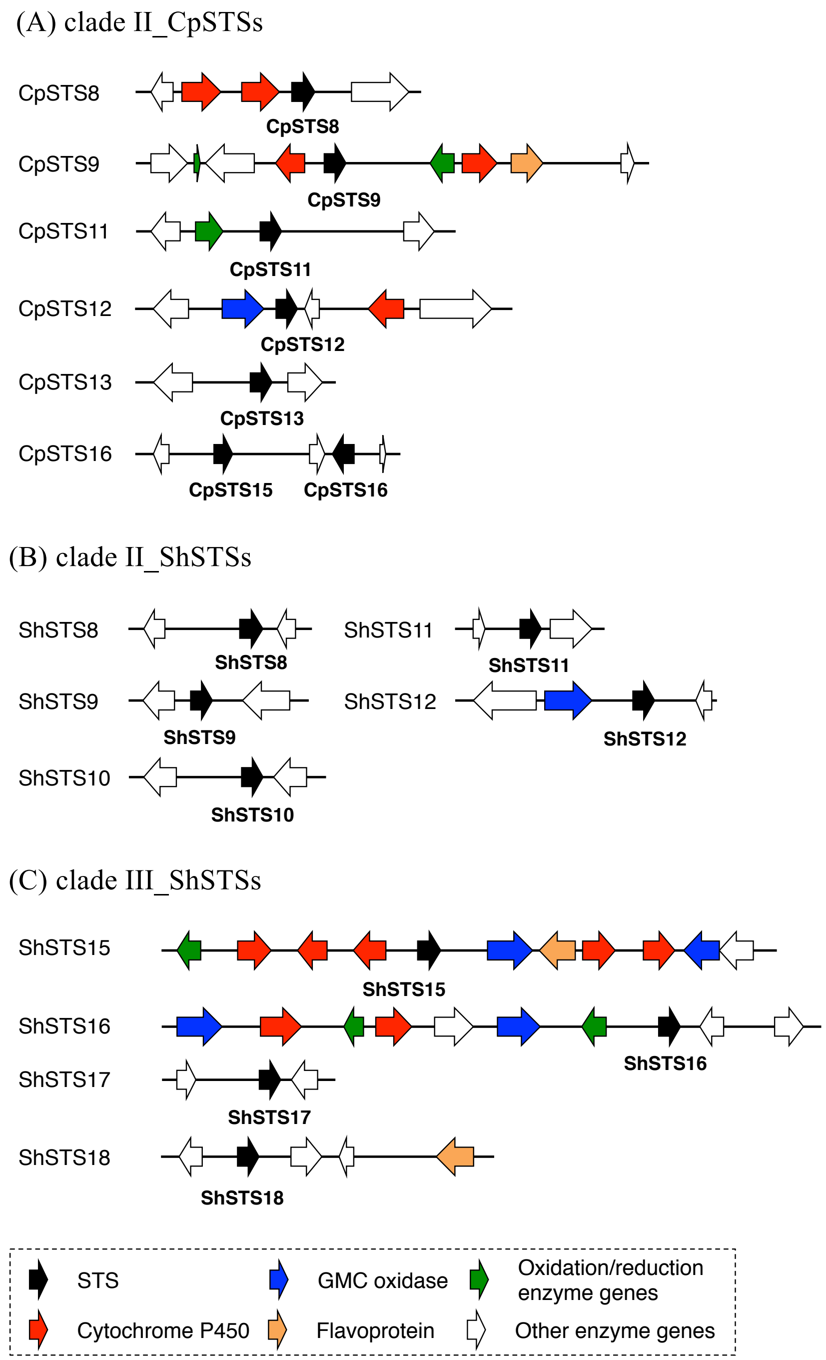
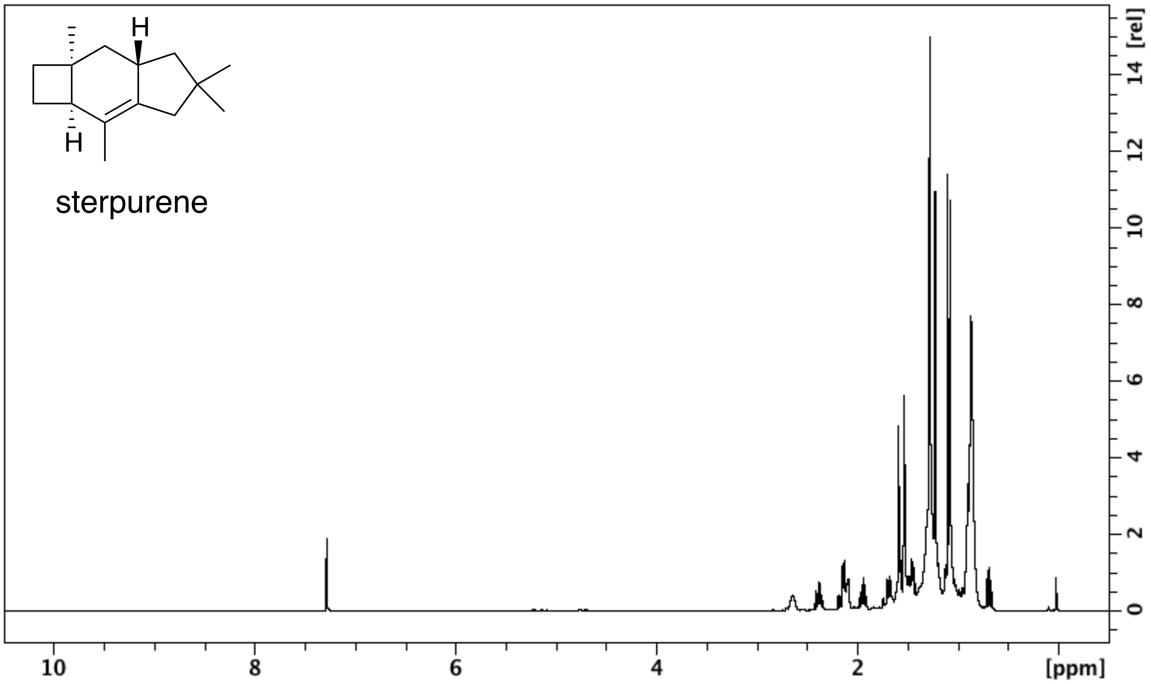


Figure S5. Schematic views of the putative biosynthetic gene clusters of (A) clade II_CpSTSs, (B) clade II_ShSTSs, and (C) clade III_ShSTSs.

(A) ^{1}H-NMR spectrum



(B) ^{1}H-NMR spectrum

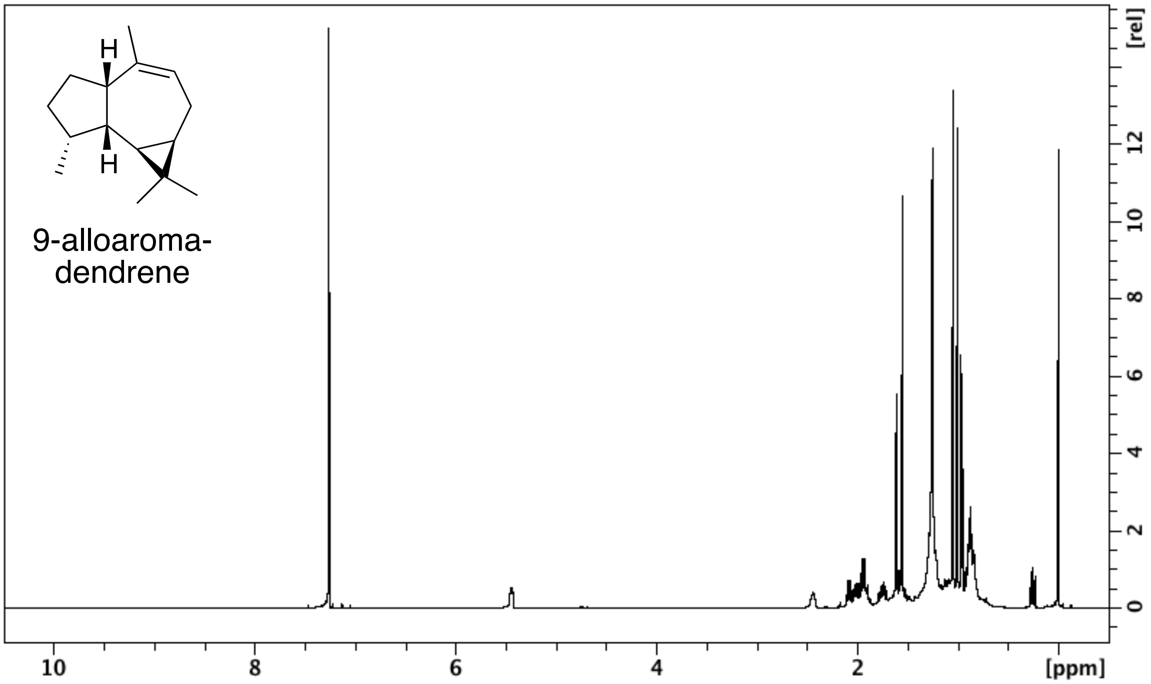


Figure S6. NMR spectra of (A) sterpurene and (B) 9-alloaromadendrene.

Table S1. Sesquiterpene synthases (STSs) found in *Clitopillus pseudo-pinsitus* and *Stereum hirsutum*. Function of ShSTS15 marked by * was not examined in this study but was reported in the previous study (*ChemBioChem* **2013**, *14*, 2480-2491.).

Gene name	Accession No.	DDxxD motif	NSE motif	Classification	Number of introns	Main product	Comment
CpSTS1	LC436345	DEYTD	NDMCSYKKE	clade III	4	sterprene	
CpSTS2	LC436346	DELSD	NDVYSYDME	clade II	4	alpha-cubebene, delta-cadinene	
CpSTS3	LC436347	DNISD	NDIFSYNVE	clade I	5	alpha-muurolene, delta-cadinol	
CpSTS4	LC436348	DEYTD	NDLCSFRNE	clade III	4	6-protoilludene	
CpSTS5	LC436349	DNSLD	NDIFSYNVE	clade I	3	gamma-muurolene, alpha-muurolene	
CpSTS6	LC436350	DEYSD	NDLYSYNVE	clade III	5	pentalenene	
CpSTS7	LC436351	DEITE	NDVFSFKVE	clade III	4	alpha-farnesene	
CpSTS8	LC436352	DEYTD	NDVYSYNME	clade II	4	alloaromadendrene	
CpSTS9	LC436353	DEYTD	NDLFSYNME	clade II	4	ledene, virifloridol	
CpSTS10	LC436354	No motif	No motif	N. E.	2	–	
CpSTS11	LC436355	DEATD	NDIHSYNME	clade II	4	9-alloaromadendrene	
CpSTS12	LC436356	DEYTD	NDLYSYNME	clade II	4	beta-elemene, ledene, virifloridol	
CpSTS13	LC436357	DETTD	NDIQSYNME	clade II	5	ledene	
CpSTS14	LC436358	DDYILE	NDIYSYKVE	clade IV	2	beta-elemene, beta-farnesene	
CpSTS15	LC436359	DDLME	NDLFSYRKE	N. E.	2	No products	
CpSTS16	LC436360	DESSD	NDIHSYNME	clade II	4	aristolene	
CpSTS17	LC436361	DDIIE	NDLFSYRVE	clade IV	2	beta-caryophyllene	
CpSTS18	LC436362	DDLSD	NDLCSFNKE	other clade	8	gamma-cadinene	

Gene name	Accession No.	DDxxD motif	NSE motif	Classification	Number of introns	Main product	Comment
ShSTS1	Stehil159379	DDSLE	NDLMSFYKE	clade IV	2	alpha-barbatene, beta-barbatene	
ShSTS2	Stehil113028			clade IV		–	62 % identity with ShSTS1
ShSTS3	Stehil122776	DDYID	NDFFSYLKE	clade IV	6	beta-farnesene, alpha-farnesene	
ShSTS4	Stehil52743	DDYID	NDFFSYLKE	clade IV	14	hirstene	
ShSTS5	Stehil161672	DDLSD	NDLCSFNKE	other clade	8	gamma-cadinene	
ShSTS6	Stehil45387			clade I		–	60 % identity with CpSTS3
ShSTS7	Stehil167646	DDWTD	NDIFSYNVE	clade I	3	delta-cadinene	
ShSTS8	Stehil146390	DEISD	NDIYSYDME	clade II	4	alpha-cubebene, 1-epi-cubenol	
ShSTS9	Stehil155443			clade II		–	68 % identity with ShSTS8
ShSTS10	Stehil111121	DEISD	NDVYSYKVE	clade II	3	germacrene D, delta-cadinene	
ShSTS11	Stehil128017	DEISD	NDVYSYNVE	clade II	3	alpha-cubebene, delta-cadinene	
ShSTS12	Stehil111127	DEISD	NDVYSYSME	clade II	3	alpha-cubebene, beta-cubebene	
ShSTS13	Stehil50042	DDILD	NDTFSYRRE	clade III	5	beta-caryophyllene	
ShSTS14	Stehil70268			clade III		–	66 % identity with ShSTS13
ShSTS15*	Stehil64702	DEHSD	NDIVSYNIE	clade III		delta-6-protoilludene	
ShSTS16	Stehil73029	DEYSD	NDIVSYNLE	clade III	4	delta-6-protoilludene	
ShSTS17	Stehil69906	DEHSD	NDIVSWNLE	clade III		delta-6-protoilludene	
ShSTS18	Stehil25180	DEYSD	NDICSYNVE	clade III	4	delta-6-protoilludene	

N. E.; not examined

Table S2. Identification of sesquiterpenes described in this study. Sesquiterpenes were identified by comparing retention times, retention indices, and mass spectra with those of reported data.

CpSTS	ShSTS	Product	retention time	RI_ observed	RI_lit. (NIST)	RI_lit. (mass finder)	Comment
CpSTS6		pentalenene	15.03	1337		1343	
CpSTS2	ShSTS8, 11, 12	alpha-cubebene	15.35	1347	1351 (HP-5)		doi:10.1016/j.jep.2006.04.018
CpSTS1		sterpurene	15.43	1350		1351	See Figure S8
	ShSTS4	hirsutene	16.14	1372		1378	
CpSTS4	ShSTS16, 17, 18	delta-6-protoilludene	16.27	1376		1382	
	ShSTS12	beta-cubebene	16.66	1389	1390 (HP-5)		doi:10.1016/j.jep.2006.04.018
CpSTS12, 14		beta-elemene	16.68	1389	1392 (HP-5)		doi:10.1016/j.jep.2006.04.018
	ShSTS1	alpha-barbatene	17.20	1406		1414	
CpSTS17	ShSTS13	caryophyllene	17.46	1415	1419 (HP-5)		doi:10.1016/j.jep.2006.04.018
CpSTS16		aristolene	17.48	1415		1423	
	ShSTS1	beta-barbatene	18.22	1439		1445	
CpSTS14	ShSTS3	beta-farnesene	18.67	1454	1454 (HP-5ms)		doi:10.2298/JSC100203125P
CpSTS8		alloaromadendrene	18.77	1457	1461 (HP-5)		doi:10.1016/j.jep.2006.04.018
CpSTS3, 5		gamma-muurolene	19.28	1473	1477 (HP-5)		doi:10.1016/j.jep.2006.04.018
	ShSTS10	germacrene D	19.33	1475	1481 (HP-5)		doi:10.1016/j.jep.2006.04.018
CpSTS11		9-alloaromadendrene	19.62	1484			See Figure S8
CpSTS9, 12, 13		ledene	19.82	1491	1495 (HP-5)		doi:10.1002/ffj.1812
CpSTS3, 5		alpha-muurolene	20.01	1497	1499 (HP-5)		doi:10.1016/j.jep.2006.04.018
CpSTS7, 14	ShSTS3	(E,E)-alpha-farnesene	20.22	1504	1509 (HP-5)		doi:10.1016/j.jep.2006.04.018
CpSTS18	ShSTS5	gamma-cadinene	20.57	1517	1524 (HP-5)		doi:10.1021/jf0303316
CpSTS2	ShSTS7, 10, 11, 12	delta-cadinene	20.68	1520	1524 (HP-5)		doi:10.1016/j.jep.2006.04.018
	ShSTS3	nerolidol	21.83	1560	1565 (HP-5)		doi:10.1002/1099-1026(200007/08)15:4<245::AID-FFJ904>3.0.CO;2-V
CpSTS9, 12		viridiflorol	22.64	1588	1590 (HP-5)		doi:10.1002/ffj.998
	ShSTS8	1-epi-cubenol	23.65	1624	1623 (HP-5)		Custer, Y., GC Volatile Components Analysis of Different Parts of Litchi chinensis (Dissertation), 2009.
CpSTS3		delta-cadinol	24.19	1643	1645		doi:10.1016/j.jep.2004.02.010

Table S3. Sequences of (A) spliced introns and (B) nonspliced introns in AO-transformant. 5'-splice and 3'-splice site are shown in red bold character. Polypyrimidine tracts were shown in purple italic character. Branch sites with CTNAN and TTNAN are shown in blue or orange bold character with underline. Introns marked by * were partially spliced in AO-transformant, thereby generating multiple cDNA sequences. Consensus sequences at the 5' splice site, the branch site, and the 3' splice site of (C) spliced introns and (D) nonspliced introns.

(A) spliced introns

Region	Sequence	Intron length (nt)	Branch site position (nt)
Group I, Branch site motif; CTNAN			
CpSTS1-intron1	TCAGTGAGTTCCTCATCGGTTTGTCTTCGAGCTGCTCCGCTAAACCCTCGTAGGTC	50	37-41
CpSTS1-intron3	AACGTATGGCTCTCGTCTACTGTTTTTCTCTGCGTGTTTCTAACGGCCAAGGAATATAGGAC	56	37-41
CpSTS1-intron4	CAGGTGAGTATCCGAAACCCTGGTTAAAGGCTTCTTAACTCATGTATCCTCTCAGATA	52	36-40
CpSTS2-intron1	TTCGTAAGTCGTTCGCATCGTATCAATGGATCGAGTTACTAATCCAGTTATGAAGGTG	52	36-40
CpSTS2-intron2	GGAGTAAGCACTCTACAGTTGTGGTACTGAACCCTCCCACTGATGAACTGGCTTAGATT	53	37-41
CpSTS2-intron3	AATGTGAGCTTTTTCTCTTTTTTAGATATGTGGATTCGATCTAACGTGCGTTCTCAGGAT	54	38-42
CpSTS2-intron4	AGAGTAAGCAGACTCCCCTCCGCAGTTGCCTGTGTAGATGATCTAATGTTTTCTTTCCGTTACGCAGTTG	64	40-44
CpSTS3-intron1	TCTGTAAGTTTTCTTCTCCCTCTTCCGTTATTTTTTTTTCTTCGTCAGCGCTAATTGTACTCTTCATTTCAGGGA	69	48-52
CpSTS3-intron2	CGAGTATGTCTTCCCGATCCCCATAGTCAAATGCTAAGCGCTGACATTATTCTTGTCTCCTAGTTT	60	38-42
CpSTS3-intron3	CTGGTAAGTTTTTACGCGATGCACCTCCTGTCTTTTCATTCCGCTGACCCTATGTTTCATAGGAT	59	41-45
CpSTS3-intron4	AATGTACGTTCATTTCCGACTACCTTCTTTGACAACTACTCACGTTCATTACAGGAC	51	36-40
CpSTS3-intron5	CGGGTATGTACCTCTTAACTCCTTGACGTTGTCACTCCAAGCTTAACCGACTCCAACCCAATAGTTC	61	39-43
CpSTS4-intron1	TTAGTGAGTACAATTTTACATACGGTATTTGTCCTTCGGGTGGCCGGCAGGGTACTGATATCTCCTTAGATT	66	52-56
CpSTS4-intron2	AAGGTAAGTCTGTTGGACCTCCTGTAACTCTGTCTCACTGACTGTCTCGGGGTAGAGC	52	35-39
CpSTS4-intron3	ACAGTATGTATCATTTAACTGTTCTAGCTGGCTGGCACTTATGTAGCAATTCGCGATAGGTT	56	35-39
CpSTS5-intron1	AGAGTATGTTCTTCCTGTTGACAAGTTCACTTCAAAATACTCATGCATCCCCTATAGTTA	54	37-41
CpSTS5-intron3	AACGTATGTACCCTATTTTCCGCCTTGTTTTGGCGCCTCATGTACTGACATCACCTTGCTTCTCAGGAC	63	42-46
CpSTS6-intron1	TTGGTACGCTCCGTTGTCTGCTCTACTGTTTCCCTCGTACTGACGGTTGCTAAGGGC	51	37-41
CpSTS6-intron2	GAGGTATGTTCATTTTATCACTAGAATACGTCAATGGATGCTTAACCATTTAACAGATG	53	38-42
CpSTS6-intron4	TGAGTGAGTGGCCTAATTGTCAGCATCACACATAGTCGGTTCTGACCAAGGTTGATAGACA	55	39-43
CpSTS6-intron5	GAGGTCAGTCTATCGAAAGCTTCGGAAAAGACATGGGACTTAACCCACAAACAGAGC	51	36-40
CpSTS7-intron1	TTGGTACAGAAAAATCTGACTTTGAGGTATTGGCTTCCAGCTGACCAAACATTGTTAGGTC	55	38-42
CpSTS7-intron2	AAGGCAAGCTAGCTTGTTATGATCCCTGAATAAGGTGCTGATGTGATACTAGAGC	49	35-39
CpSTS7-intron3	GCAGTATGTTCAGTCTTTTCAATTTCTCTACCATGACTGAATGTTCTCAGGTT	47	34-38
CpSTS8-intron2	AGAGTAGGTGCTGAGAATTATCGGACCCCACAACATCGGAAGGCTAATGCAATGAATAACATCTAGGTT	63	41-45
CpSTS8-intron3*	AATGTAAGTACACATCACAGAATGATATCATGATTGGCTGAGCTTGTGTGAAGGAC	50	35-39
CpSTS9-intron1	CTAGTAAGCGCCTATTTTACCAAACTGTAAACAGGCTTCACCACTGACCGTACTTTCCAGACA	57	41-45
CpSTS9-intron2	AGAGTCAGTTCTAGTGGCACCTCGAAAATTTCCTATCTAACGTTTTCGATAGTTT	49	34-38
CpSTS9-intron3	AATGTAAGTCAGGCTAGGAAAATTCTCAAGGGACAATGAGCTCACTCGTTGCTGTAGGAC	54	38-42
CpSTS10-intron1	CTGGTCCGTGAACGCTCAATAATACCATTTTCCTGAACTAACATCAGTGTTCAGCTT	51	35-39
CpSTS10-intron2	GTAGTAAGACTTCGTCCAGCAATGTTCTGAGTTCGGCCTTGCAAACTAAACCATTATCTATCACACAGGAT	65	43-47
CpSTS11-intron1	CTGGTAAGAGATGTGATCACGTTCTGCATGCCTCGTTTTCTAACTAAGCTCGCAGCAC	52	37-41
CpSTS11-intron2	AGAGTATGTGAAATCACATACGTCAATTGGGATCGAGCTAATCAGGTACTGGCGTCGCGTTCAGCTT	61	35-39
CpSTS11-intron3	AACGTGAGTCCTCATCCATCTTCTTGACTAATGTCTGATAGCCATTCACTCGTAGGAC	52	32-36
CpSTS11-intron4	CAGGTGTGTTTTCCATCTTACGTACCGATGATCCCTATTCTGACTTGAACAAAAATTAGGTG	56	37-41
CpSTS12-intron1*	CTAGTAAGTAGAGTTCGTACGGTCGTAGCTCGGTTGGATCTGAACTTCCACCTGCCAGGTC	55	37-41
CpSTS12-intron2	AGAGTGAGCGTCCTTCCTCACGTCCGAGTGCTAGACTCTTAGATTTATTTCTTTAGCTT	53	35-39
CpSTS12-intron3	AATGTGAGCCTGGAGAAAGGTTCTAATTCGTCCGTTGCTTACCGTCCGTCGCAGGAT	51	35-39
CpSTS12-intron4	AGAGTATGTGATCAATCGGAGCATACACTGGCTGCCTTGCTGATGTCGAGCCACACAGATG	55	37-41
CpSTS13-intron1	TTGGTAAAAGATCTCCTCTAATATACCTACGTTGTCCGCAAGTATGCTCATTCGAAACTATAGTAC	60	44-48
CpSTS13-intron3	AATGTGAGGACTTCCTTCGATAACCCGGGAACTGAGGCGGAAGTGTTTGACGACTAATCCGTTCTTGGCAGGAT	68	51-55
CpSTS13-intron4	CAGGCAAGTGTTCCTGATTCATTTTCTTCCGCGTTGCTAATTTGAGGCTCACAGATG	51	34-38
CpSTS14-intron1	AAGGTGAGTGGAATGGTTGGAAGTGGAGGGTGTAGAGGAAGAAACCGAGGCTGACGTTGGTATCATAGAAG	65	48-52
CpSTS14-intron2*	CTTGTGCGTACTCTATGCCTTCTAGCATTCATCCTAGTCTCGACCTGATGATTTACAACCCTTCAGCCT	63	42-46
CpSTS15-intron1	GAGGTACGTGGGTTGATGACATTGGGAGTCGGAAGATCTCACTGACCTCGTCCGTGAGCAGCTT	58	39-43
CpSTS15-intron2	TTGGTATGACGTCCCAGTTGATTGATCAACATGACTATCACTGACCTCTCACATTTTCTGTTAGGCT	61	38-42
CpSTS16-intron2	GGAGTGGGTTCTTTCGTTTCATTCATTTTTTCGTCCAACGTTCGTACTGAAACGCAGTGGCACAGTTT	62	44-48
CpSTS16-intron3	AACGTAGGTGGACCTTGAACACTCCTCGTTAAGTTGCTAACGCCCATTCCTTCCATTTCGTGATCCAGGAT	65	34-38
CpSTS16-intron4	TATGTAAGTTCTCGACACTCGTTCATCCGGAAGCGTCTGCCGGTTCTAAACTTTTTCTGTATTCATAGATG	65	43-47
CpSTS18-intron1	ACAGTAAGTCTAATCAAGCTTCTAATTAAACGGGTCGCTAAGTCGTCACCGTCAGATT	52	44-48
CpSTS18-intron2	GCTGTTCGTGACCTCTAAATTCATGTCCCCATTTGTTTTACTGACCACGCGGTAGACG	52	38-42
CpSTS18-intron3	GGAGTGAGTAACATTTTTCTCGGAAGAAATTTCAAGAACTCAACATTTCAATTTAATTCAGCTT	58	36-40
CpSTS18-intron4	ACGGTGAGCAAAATCACGAGATTTTTCGACGAATTGAGATACTGACTACCTCCCCTGTTTTCGTAGGTT	63	39-43
CpSTS18-intron5	AAGGTACTTCAGTCAATTCATTCGGTTCGTTCTCCGGACTAATACCCTCGATACAGCCA	53	36-40
CpSTS18-intron6	AACGTGAGTTATTCCGGACATCATTGAAACCTTCGCTAATCAACTCAAATTATCTCAGGAC	55	41-45
CpSTS18-intron7	AACGTGAGTATCCATTATTAATTCGCTTAAAAAGCTTTTCTGACAGGCTGTGCAGAAG	52	37-41
CpSTS18-intron8	CTCGTAAGTTACCGTATCTTCCTCGGTCCAGGGTCTCTTTACTGACATTCGTGGTATTCAGGGT	58	39-43
ShSTS1-intron1	CTTGTGAGTTTCCGAGTTTGCATGACATGCTGGTCTTGTGACTGACCTGATCTTTGTGCTCGCAGTCC	62	39-43
ShSTS1-intron2*	ATCGTCCGTGCACTTACCCTCTTCGCAACGAGCATTGCTCATCATTCGTGTTTTACAGGAG	55	35-39
ShSTS3-intron1	GAGGTATGTCGTATGATTGTCTATTGTCCATACCCTTTGCATCCTCATACTGACTCCACGTCCTTCGCAGTCG	67	41-45
ShSTS3-intron2	CCTGTACGTGGTCTACCCCTTTCCACTGTTATCCTCGGTATCACGAGAACTGACTGAGCACTTCCCCATTATTATTTCAGCCC	77	47-51
ShSTS3-intron4	CAGGTAAGCAACAGAACTGAACGCTATGATAGCTCACTAGCTGATCACCGTCAACAGAGG	54	38-42
ShSTS3-intron5	ATGGTGAGTCATCGTATCCATTGATATGATCCATAACTGAACATTGCTTTTGCAGAGG	52	34-38
ShSTS4-intron1	AAGGTATGTTCATACTCGTTATTGGCTGCTGTGTTTTTGCTTACGCCACTGCCTCAGAAC	54	37-41
ShSTS4-intron2	GGTGTACGTATACAATGTGTCTCTGCTTCTACAATCCTCTAACCAGACGCATAGGTG	51	36-40
ShSTS4-intron3	GCGGTATGCACGTTTCTTGTATTGACTGCAGCTACTCGAGCCTGACGCACAATGCCGTACCTAGGCT	61	39-43
ShSTS4-intron6	ATGGTCAGTCTATGCGTTTACTACCGATTCGTCACGTTGCTAATATCGTCTATCCCGTTCTAGTAT	60	37-41
ShSTS4-intron7	ACGGTAAGAACCAAATATAAAGACCGTGCGGCAAATTCTAATAGCTATCATTTCAGCTT	53	35-39
ShSTS4-intron9	GAGGTAAGCGAAAATCACTATGGTATCCACGTACTCGACATCACCTAACAGTTCGTTTACAGAGC	59	42-46
ShSTS4-intron10	CGCGTAGGTTGCTCGGTTTATTCTTCTCTACGCAATATTCCTAATTACTGGTACAGTGC	53	38-42
ShSTS4-intron11	AACGTATGTCCTTTTTTGTTCTTATTGAGCCTATTGAGACCACCTAGCTGACGCACATTATTTTAGCTA	63	45-49
ShSTS4-intron12	GGCGTACGTTATACTCACTCTTCTTATGGCAGACCACTCAGCTGACAGATATGTTTCAGTCT	56	39-43
ShSTS4-intron13	CTGGTACGCAGTTCCCCTTCGCGAGTGTTTGTTTGATTTCTCAACCTTTTTATTCGCAGACC	56	37-41
ShSTS4-intron14	AAGGTCAGTTGCTTTCATCTTACCTGTTCCTCTTCCACCTACTGATTGATTCGCTCTGCTCTCTAGCTT	63	39-43
ShSTS5-intron4	AAGGTGCATACTAATCTTATTGTCAATAAACTACGCTCACTCGACCACTAGGTT	48	33-37
ShSTS5-intron5	AAGGTTCGCAGTCGTTCGCTTCATCCTTTAGTATCAATTCTGAGATCCAGATTCTAGCCA	54	37-41
ShSTS5-intron6	AACGTACGTGTCCTGCAACCATCTCGAAGGATTACCATCTCACTGTACCGGCCGTGATTAGGAT	58	36-40
ShSTS5-intron7	AATGTTCGTTTTTCTGTGTTTATTACCCCCGCGTGGTGTACTCATCGTCAAATAGAAG	52	38-42

Table S3. Continued.

ShSTS5-intron8	CTAGTATGTCGCCCTTACCAGTAAACAGCCCATTGTGCGACGACTCATTACTTTGGAAAGGAT	57	41-45
ShSTS7-intron1	GTCGTGCGTGCAATTTATTATTAGCCCTCTATTCCCTAGTCTCTGAGTCATCATACTTGGTTGCATCGACAACTGTACTGATGACTGCCTTCCTATTCTCAGGTT	99	75-79
ShSTS7-intron2	AATGTGAGTAATCCAATGCCATCGTATACCCGGCATCCTCTTCAGGTCTGCAAGTCAAGGAGCATAGCTGACCCACTTCGTGTCCAGGAT	84	65-69
ShSTS7-intron3	ACGGTCGGTTCCTCCATGGCGTCGTTGGTTACGTGATCGAAGCTGATGCCATTGCCTGCAGGTA	58	40-44
ShSTS8-intron1	AGAGTACGTTCATTCAGATCCGCGCAAACTTGATAGGCTATTCTGTCCCGATCGACTGACGCCCTACCTGTTCTGTCTAGCCT	77	53-57
ShSTS8-intron2	CCCGTAAGTCTGCGCGTTACCGTCCAACCCGACGACCTGATATCTGATCTTCACAAAGATA	55	41-45
ShSTS8-intron3	AATGTAAGCCTTGCCTGCTGTATGCCATGGCTCACACTCATCGACTAATTGTCACGCTAACAGGAT	60	42-46
ShSTS8-intron4	TAGGTAAGCATTGACAGCCACAGCTCTGGGGTCTCCTGCCCTTCTAACATTTGTTGGCTTTGATCGTTACAGGTA	69	41-45
ShSTS10-intron1*	GGAGTGAGCACTATTTTGCTTCCGCATTGGACTGTCATTTAAGCGTATCCTTTCTGACCCTCTACTATTGGTATTATTAAAAGCTT	80	51-55
ShSTS10-intron2*	AACGTGAGTAGTTATATCTCTCCGGGATCTGGCGACTAATTGATGGTTCCATTCGATAGGAT	56	33-37
ShSTS11-intron1*	ACAGTAAGCACTAGCGTGTTTTTTTTTTCGCTTCTTAAGCTCAGGGCGCACTGACGCTAACATGCACTTTGACTTTGGCTCAACTCATTATCCGCAGATT	94	76-80
ShSTS11-intron2	AATGTGAGTGACACCACACACACACCTCACTCCCTCCCTTCCTCTCCTATCAATCTTAGGCACAATGGATGTCGGGGCTGACATATGTCATAACTTGGTCGGAAGGAC	102	75-79
ShSTS12-intron2	AATGTAAATATGTTCCCTCCTCCCGGTTATCTCATTCAGGTGTGACGGTTACCCAGGCGCAATCTCAACTGACGCTCTCTCATTGTAGGAT	85	66-70
ShSTS12-intron3	GAGGTGGGTAGCTGTCTTCGACAAATGCGTAGTTGCTCATATCCTTTTTGTTTAGATA	52	33-37
ShSTS13-intron2*	GAAGTAAGGAGATCATTTATCCCTGCGTACTTACACCCGTAGCTGAATTGTTATCATGCCAAGCTC	60	40-44
ShSTS13-intron3*	AAGGTGAACATTAACAGCAGTGGTCAAACGGCCTCGGTACTCACACTCTCCGTGATTATCAGAAC	59	37-41
ShSTS13-intron4	CCAGTCAGTGTTTCAAGCTTCCTTCGTACGGTTGTCTAGCACTGACTCACCAGAAATATCGCACTAGGTT	64	39-43
ShSTS13-intron5	AACGTGAGTTGAACCAACACCTCGTTTTCGTCTCACATCGCTGACAGTTCAGGCAGGAC	53	38-42
ShSTS16-intron1	TCAGTAAGTCTTTCATAAATTTCCGCAGCTTATTCTTCGCCTATTTGCTTACTGTCTACGTTTTTGTTCGCGCTTTTGCTCAGACC	80	45-49
ShSTS16-intron3	GCAGTATGTCTCCCTATTTCTTCTTCTTCGCCGTGCTCCAGTTATCCTGTCCTTCTCCTTCCTTGTATTATTATCGTTTGTTCTTCATGATCCCGGCCATGCTCATCACTCATTTTTATTTTTAGATT	122	106-110
ShSTS16-intron4	AACGTAAGTAACAATGCTGAGATACTCTGCACTCACCGTGCTAACCCGATGTTCTTGCTTAATAGGAC	62	38-42
ShSTS17-intron1	TCAGTAAGTTCCTTTTGGGCCGGCAATCTAGTAATCAAGCATGTCTCACCAGTCATACTGCGTCGGTAATATAGACC	71	42-46
ShSTS17-intron3	ACAGTGAGTGTAGCCTGCAACAGTGCCACATGATATATTCCCAATGGGTCGTGGCTCACTGAGAAGAACAGGTA	68	52-56
ShSTS18-intron1*	TCAGTGAGTTTACAGTCAATTCCTATGTCGAGGACATCGAGTGGGCCACTCACCATAGTCTACTATAGACT	65	46-50
ShSTS18-intron3	ACAGTGCGTTCCACTGCAAACCCACACCAAGAAAGTACTCTGACACGGCAGGGACAGGTT	54	37-41
ShSTS18-intron4	AACGTACGTTAATCTGACCACTTCGCAGAGACAGCTCATTCAGATACTCTGGACCGTACAGGAC	58	32-36
Ple1-intron 1	ACAGTACGTCAACTTTCCCCTTGTATTATGTGCAGAAACTCATTGAATCCCTTGGAAATTAATGGAAAATAGATT	69	36-40
Ple1-intron 2	CAGGTATGGTTCCGACATGCTGCTATCGATTTCTGTCTGACACGTTCTACTAGGCC	50	34-38
Ple1-intron 3	ACTGTGAGTACAGCGTTCTCTATCAAACGGGGCTGAAGACCGACATCACGTTATCAGGTC	54	30-34
Ple1-intron 4	GAAGTATGTATTCAGGCACACCTTCTCATACGCCATAGACCTAATCTCTTTGATAGTCT	53	38-42
Ple1-intron 5	ATAGTAAGCAGTATCTTCTAACTCTGACTTCACTTACGCCTAATCCGTCTATCCTTAGAAA	55	37-41
Ple1-intron 7	CCGGTATGCTATCTATCCCTCGAGCAGAGGACTTAGATGGGCTCATGTCTGTTATGTAGCTG	56	39-43
Ple1-intron 9	AAGGCTTCGGCAGAAGGTCAAAAACCGTTCTTCAAACCACCGTAATCATGCTCACCTTTCCTTTTAGAAT	64	48-52
Ple1-intron 10	ATCGTGAGTAGTTCGCCGATGAACATTCGGCCTTAGCTAAACCCTGTCTACAGACG	50	34-38
Ple6-intron 1	TCAGTAAGTCGCCCAAAGCTCATCCTTGATGATGGCTGACATACTCTCTAGGCT	48	33-37
Ple6-intron 2	AGCGTGAGTTGTAACCATCTATTGGATTCGAAAGGACTGCCCTGAGCCTATTCCATACAGTTG	57	39-43
Ple6-intron 3	ACAGTATGACAAGCCCTTCGTATCCGGCTGTTTGTCCAACTGACTCTGGCCTAGAAT	51	37-41
Ple6-intron 4	AAGGTACGGTATTTCGATCACCATCACTAGCCCTTCTGACAAAGACACTTTACAGGAA	52	33-37
Ple6-intron 5	TGCGTTAGTTTCTCCTTTTTCACCGTCACGGCATGGTCAGGAATCAATACTGACTTATATTCTGCAGCTA	64	47-51
Ple6-intron 6	GAGGTTTGTGGATTTGTTCGAGATGATGCATTGGGAGTGAACTCAGTCATCGTTGCAGGTA	55	39-43
Ple6-intron 7	GAGGTAATTGACTTTGTTCTTCGTTAGAATTGATAGAGCTAACCAAACTCTAGTAC	50	36-40
Ple6-intron 8	TTTGTAAGTCCTGCTCTTCTGACTTTCGTTTGGCATACTCAGAACTTGGAATAGATG	51	35-39
Ple6-intron 9	CAAGTTAGTAGCCTCCTGCTTGCTAATCGTAGAACCATACTGACCTCTTCGCAGAGG	51	37-41
Ple6-intron 10	AAGGTAAATCTTACTCATGGCCTTAATGGCAATTGATAGCTCATGTTCTGATAGCGC	51	37-41
Ple6-intron 11	CCCGTGAGTGCTATACAACCTGTATCAGAACAATTTTCTCAATTGTCTACTACAGGGT	52	35-39
Ple5-intron 1	GCAGTATGTTTCGCCTGCACTCAACTCGGCCGCTCTTAGCTAATTGTTACCTCAGATT	52	37-41
Ple5-intron 2*	TCTGTAAGTAGTGCGAAAACTATGGAAGCAATAGGTTCCTCATGTTCCCTAGCTC	49	36-40
Ple5-intron 5	CCTGTCAGTTCGACCTATCTTCCCAATTACGCGTGTTCTTAACTGACTTACCGACGATAGAAC	57	40-44
Ple5-intron 6	ATCGTACGTAAACCTCTCTGGGGACAAGTTGATTGAAAATCTAACAACCATTTCCAGGAA	54	38-42
Ple5-intron 7	GAGGTGGGTCTCACGGAAAGACATCTGTCAGTTCTATTGTTCTCAATCGGGTGGCTTAGGTT	56	39-43
Ple5-intron 8	TCTGTGAGTTCACGAACGTGGTATATTATCGCCATTGACGGACGCTGACGGGCCTCCACAGTAT	58	42-46
Ple5-intron 9	TCTGTGAGTTGACCACCTCGCGCCATTCGGAACGTTGAACTGATTAGTGTCTAGGGG	51	37-41
Ple5-intron 10	CTGGTACGTTTCTGATCACAACTCGATTGAGTGCTTGCTTACCGACTAGAGAACCCTACAGAAC	58	35-39
Ple5-intron 11*	CGTGTGAGTGTCATCGTGTTCTTCACTCGGTACCTGATCATTTACGTCTGCTCAGAGA	52	31-35
Ple5-intron 12	ACTGTACGTTATGTTCATACGTACGGGTGGAAGAAATTTTGCTGATATGAACTGGCACAGGGT	57	39-43
Ple5-intron 13	TCGGTGCGTATAGTCGCTTATCATTTCTCAAGACACGGCCCCCGAGGCTAACCATCACTCTTTATTCTGACAGGTT	70	45-49
Ple3-intron 1	TGGGTATGTTCTTTCAGGGTCCATTCGCCGATTCAGTGCTCATTCTGCGTTTCAGTCG	52	36-40
Ple3-intron 2	TCGGTCAGTCCTTCCTCCAGCATCTCGATATATCATAACTCACAAACGACCTGTTTAGTTC	55	36-40
Ple4-intron 2	ATTGTGAGCATTCTGCTGGGATAATATTTCTACTCGAGATATCTGACGCACTTGATAGAAT	55	40-44
Ple4-intron 3	AGAGTGGGTTTTTCCACCAGTTCCTGTCATTCACGGACATATCCTTACCGACCCCTTACTAGAAA	59	41-45
Ple4-intron 4	GTGGTATGCTCCAGGCAATCGCTTGGTCTCGATAATTAACTGACAGCTGGGGATCTCCCTTAGAAA	60	37-41
Group II, Branch site motif; TTNAN			
CpSTS1-intron2	GCAGTAAGTCAATCTTTGCTATACTGGGGGAAAATGAATTGATGGGTTCAAAGATT	50	36-40
CpSTS4-intron4	AATGTGAGTCGATCATCCGATATGGGGATTGCGAAAGATATTTAACAAAGTGTGCAGGAT	54	38-42
CpSTS9-intron4	CGAGTAAGTTCACTTTACTCCTATCGAGGAAGAGGTATTAACCCCTGGAACAAGTTG	51	35-39
ShSTS4-intron8	CCGGTAAGAGTCTTCCGCCCATTTCAATTGATGCTGTCGTCATTAAATATCACCTCTCAGACC	57	40-44
ShSTS5-intron1*	AAGGTCTGCACGGTGAATTTTTTGATTATGACACAATTTGACACCACCTACTTCACAGTTT	55	35-39
ShSTS5-intron2	TCGGTGAGTGTCTGTGACCATCTGGAAGGCTGGTCTCGTCGATAACCGCTGCTTCTATAGACC	57	39-43
ShSTS10-intron3	GAGGTGAGTGATTTGTCCACTTTCGTCGTGTGAGACTGCAATTTTTGACGTTTCGTTGTGTGCATAGATA	64	42-46
ShSTS13-intron1	GAGGTTCGTCCACTGCCCACATATTACCGCGAAACAGCTCTGTTAACATTTCAACTCTGGACGATAGGAC	64	40-44
ShSTS17-intron2	AACGTAGGTGTATTTTCTTTTTCATTTGTTAGGCTTACTCTTCGTGTTGAAGTATTAATTTGGGATTTAAGAGC	68	44-48
ShSTS17-intron4	AATGTAAGTGATATTGCGCCTTTACCGTCCGTCATTCTGCCAGTTAAATATTCTGCAAAAAGGAT	59	41-45
ShSTS18-intron2	AAGGTCAGTAACTCACTCGTCCGGCTTTGCAACGACGAGAGATTTAAAGGGTGTGTAGAAC	55	41-45
Ple1-intron 8	ATGGTATGTTTGCCTTGGATCCTGTACAATTAGTCTTGACCTGTGTCTAGGGC	47	33-37
Ple3-intron 3*	TGGGTAAGTGATAGTCAATCATTGATGCTGCTGTCAAATTAAACTATCCTTTCAGAAC	52	36-40
Ple5-intron 3	GAGGTTGGTTTCTGACGGTATGCTTTGCTCGGACCGAAATTAACATGTCTATTTCTCAGACC	56	37-41

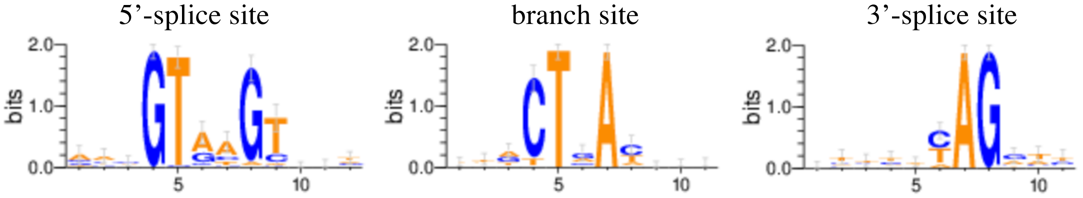
Table S3. Continued.

Group III, Branch site motif; not determined			
CpSTS6-intron3	ACAGTATGT*TTTT*A*CGTCTTC*ATTAATGATCGGGATGAAAGTTAATTTGACAACTGTAGGTT	56	38-42
ShSTS11-intron3	CAGGTAAGCTGAAGGATTGCTGAATCTAAGCATATGAAGTTGACGA*TTTCT*GGTATATGTAGGTA	59	–
ShSTS16-intron2*	AACGTATGA*TTCGCTGTCTCT*A*CCCTC*GGTGATTATACAGTAGGACACTAGTA*TTT*A*CCC*AGA*TTTTTCTCCC* CAGATG	73	–
Ple4-intron 1	GATGTAAGCA*CCTTCGTCTCTGCC*A*CTTC*ATCA*CTGTCC*GAATGACTTGGCAAAAATATAGGAA	58	–

(B) nonspliced introns

Region	Sequence	Intron length (nt)	Branch site position (nt)
Group I, Branch site motif; CTNAN			
CpSTS8-intron1	ATTGTGCGT*TCCTTTCC*AATGGTACCATGAACATAACTGAACTGCGTACTTTCAAGGGC	53	34-38
CpSTS13-intron2	AGAGTAGGTTGGACACAACA*TCTCTC*AGATTAGACTGGTTACACTGAAA*CCT*ATATTACTTGCAGTTT	62	41-45
ShSTS3-intron3	TCGGTACCCTGCCAAA*CTGCC*ACATGA*TGC*A*TTGTGT*A*CGCC*ACTGAGA*CGCCCGTTTGTGTGTTT*GACAGA GT	68	41-45
ShSTS4-intron4	GGGGTGTGTGG*TTTTTCCGT*A*TCT*ATAG*TCTTTTC*AAG*CCTT*CTGAAATATCATA*TCTCTTCTAG*CCC	62	40-44
Ple1-intron 6	TCGGTATGTGTTCGCA*TTCTTTTTC*AACATGGAAGATCCTCAGTA*TCTCCT*ATTTAGCTG	54	36-40
Ple5-intron 4	TGGGTCTGTCA*TTGCTCTGC*ACGG*TTTC*AAACA*TTTC*CTGATAACGCA*CTTTTTTTAG*ATA	55	35-39
Group II, Branch site motif; TTNAN			
CpSTS5-intron2	AATGTGCGTTA*CTTCGCCTTCT*A*CCCCTTTTC*GA*CCTC*A*CCC*GTTCAT*TTC*A*TTTCCT*AACAGACG	60	41-45
CpSTS7-intron4	AATGTGGGTA*TCTT*AGGGAGTCAGG*CCCC*A*TCCTCGCC*AATCCGTTGATA*TCCTCTTTT*GACAGGAT	61	42-46
CpSTS8-intron4	CAGGTAAGTAG*TTCC*AGG*TTC*GACGG*CTTC*AAGTATACTTGATCAA*TTTT*ATGATGGTAGATG	57	36-40
CpSTS17-intron1	AAGGTGCTTGTGA*TTC*ATA*CCCTTTT*A*TTC*AGTCAATTGACGAGAACTCAAAGGCT	50	34-38
ShSTS3-intron6	CATGTATGTTTGAA*CCC*A*TCT*A*CTCCCGCTCT*A*CTGCTT*GTTGATG*CGCTGCCTCTTGCT*GGTAAAGTCT	64	38-42
ShSTS12-intron1	GGAGTGAGTAAA*TCT*AACAATAA*CCTTTCTCCT*AGAGA*CTTC*GGTGACA*CTGT*AA*TTC*A*TCTTGCC*TTCATCC ACA*TGTGCC*ACCCACAGCTT	87	64-68
Group III, Branch site motif; not determined			
ShSTS5-intron3	GAGGTTCGTG*CTTT*A*CGTTGTCT*A*TGTC*ATA*TGT*ACCACGTAAGATCGGTATATCACTTCAGCCT	59	37-41
CpSTS16-intron1	CTGGTTCGTCGCA*TCTGCTGTTCGTGTTCTCTT*GGTAATGGA*CTTC*GAACAGCAT	49	32-36
CpSTS17-intron2	CTGGTGTGT*CTTCTTCGTCTCCTGTGTTTCT*ATCGAGCCAACCA*TGTTCTTTGTCCCAACAG*GTT	59	–
ShSTS4-intron5	GGAGTACGTCCATA*CGTTCTTC*ATATA*CTTT*GA*CTTCCTC*GATAACCCCTTGGCCGAACAACAGTTC	61	–

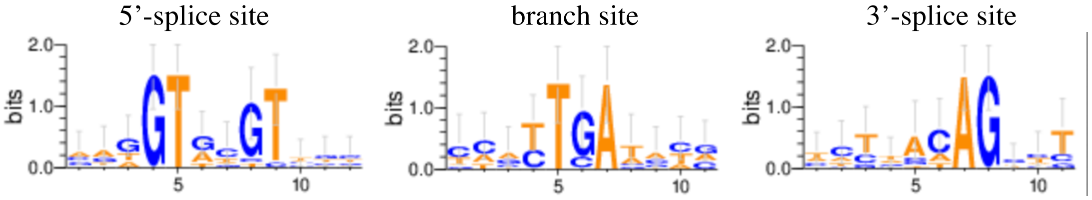
(C) spliced introns


5'-splice site

branch site

3'-splice site

(D) nonspliced introns


5'-splice site

branch site

3'-splice site

Table S4. Summary of amino acid sequence identities of STSs in the same clades. (A) clade I; (B) clade II; and (C) clade III.

(A) clade I

	Omp1	Omp2	Omp3	Cop2	Cop3	CpSTS 3	CpSTS 5	ShSTS 7	ShSTS 6	PpSTS -01	PpSTS -03	PpSTS -06
Omp1	-	56%	42%	44%	43%	41%	55%	46%	39%	50%	41%	40%
Omp2		-	52%	45%	55%	54%	75%	48%	49%	71%	49%	44%
Omp3			-	47%	73%	78%	52%	52%	61%	54%	46%	50%
Cop2				-	46%	46%	48%	52%	44%	50%	47%	40%
Cop3					-	77%	55%	50%	61%	53%	46%	46%
CpSTS3						-	52%	52%	60%	54%	46%	51%
CpSTS5							-	48%	49%	69%	48%	44%
ShSTS7								-	46%	54%	48%	45%
ShSTS6									-	51%	43%	48%
PpSTS-01										-	47%	45%
PpSTS-03											-	40%
PpSTS-06												-

(B) clade II

	Omp4	Omp5	Cop4	CpSTS 2	CpSTS 8	CpSTS 9	CpSTS 11	CpSTS 12	CpSTS 13	CpSTS 16	ShSTS 8	ShSTS 10	ShSTS 11	ShSTS 12	ShSTS 9	PpSTS -10
Omp4	-	66%	60%	59%	41%	45%	39%	41%	37%	37%	52%	52%	61%	53%	50%	68%
Omp5		-	51%	49%	38%	40%	36%	37%	34%	34%	46%	49%	52%	47%	46%	54%
Cop4			-	60%	39%	40%	36%	39%	36%	37%	53%	50%	55%	53%	51%	61%
CpSTS2				-	38%	38%	35%	36%	36%	37%	51%	47%	53%	49%	50%	63%
CpSTS8					-	56%	43%	50%	41%	44%	38%	37%	38%	36%	40%	40%
CpSTS9						-	51%	62%	48%	48%	37%	36%	38%	35%	38%	41%
CpSTS 11							-	45%	67%	51%	35%	34%	36%	36%	35%	38%
CpSTS 12								-	43%	42%	35%	35%	36%	36%	35%	38%
CpSTS 13									-	50%	34%	34%	33%	34%	33%	37%
CpSTS16										-	34%	36%	36%	36%	34%	41%
ShSTS8											-	51%	53%	50%	68%	55%
ShSTS 10												-	66%	65%	48%	55%
ShSTS 11													-	71%	53%	67%
ShSTS 12														-	49%	60%
ShSTS9															-	54%
PpSTS-10																-

(C) clade III

	Omp6	Omp7	CpSTS1	CpSTS4	CpSTS6	CpSTS7	ShSTS13	ShSTS16	ShSTS17	ShSTS18	ShSTS15	ShSTS14	PpSTS-08	PpSTS-14
Omp6	-	62%	45%	53%	56%	44%	36%	57%	55%	61%	57%	37%	51%	52%
Omp7		-	46%	55%	55%	44%	39%	62%	59%	63%	57%	40%	53%	52%
CpSTS1			-	46%	42%	39%	34%	46%	42%	43%	40%	33%	45%	43%
CpSTS4				-	48%	58%	39%	56%	48%	53%	54%	37%	55%	49%
CpSTS6					-	42%	33%	54%	52%	54%	50%	35%	47%	47%
CpSTS7						-	37%	44%	41%	44%	45%	34%	48%	42%
ShSTS13							-	37%	40%	36%	39%	65%	35%	35%
ShSTS16								-	61%	60%	63%	41%	53%	51%
ShSTS17									-	59%	54%	41%	48%	50%
ShSTS18										-	56%	37%	48%	51%
ShSTS15											-	37%	51%	50%
ShSTS14												-	38%	39%
PpSTS-08													-	60%
PpSTS-14														-

Table S5. Comparison of active sesquiterpene synthases sequences and predicted protein sequences by Augustus. Conserved motifs (DDxxD and NSE) are described in red bold characters. Sequences described in blue bold characters with underline show differences between active protein sequences and predicted one. Sequences described in orange italic characters show the missing *N*- and *C*-terminal sequences due to the misprediction. If the predicted sequences are identical to the active one, those sequences are described in grey characters.

Gene name	Active protein sequence	Predicted sequence by Augustus (*A. oryzae*)	Predicted sequence by Augustus (*C. cinereus*)
CpSTS1	MPAAIPKFYTLPDTLRNWPWKREISPYYRQCQAESVAWLESFRPFSPKAQVAFNKCDFSLVSALCFPKGSPYNLRSVCDLMHTFFTL**DEYTD**YLDLEGVKTLCEATMDAIRNPDKPRPEGEHFIGEVARQFWARARVNATPACEERFVKSWRTYLNSVIQQAERRDAKYICTMEEYLHARRDNIGSDPSFALLEITLEVDLPHEVMEHPTIVALARDTTDMIVLA**NDMCSYKKE**IIADDANYNAVTVVMHNHNTNVDGGIQWISDYHDTIVDHFLRLREDVRLKQNGFPSWGARIDREVEAYVEGLGLWIRGHDEWNFGSGRYFGDEGLEVQKSRIVECTVSADPFTPFLIQQEEEVDENAA	MPAAIPKFYTLPDTLRNWPWKREISPYYRQCQAESVAWLESFRPFSPKAQVAFNKCDFSLVSALCFPKGSPYNLRSVCDLMHTFFTLDEYTDYLDLEGVKTLCEATMDAIRNPDKPRPEGEHFIGEVARQFWARARVNATPACEERFVKSWRTYLNSVIQQAERRDAKYICTMEEYLHARRDNIGSDPSFALLEITLEVDLPHEVMEHPTIVALARDTTDMIVLANDMCSYKKEIIADDANYNAVTVVMHNHNTNVDGGIQWISDYHDTIVDHFLRLREDVRLKQNGFPSWGARIDREVEAYVEGLGLWIRGHDEWNFGSGRYFGDEGLEVQKSRIVECTVSADPFTPFLIQQEEEVDENAA	MPAAIPKFYTLPDTLRNWPWKREISPYYRQCQAESVAWLESFRPFSPKAQVAFNKCDFSLVSALCFPKGSPYNLRSVCDLMHTFFTLDEYTDYLDLEGVKTLCEATMDAIRNPDKPRPEGEHFIGEVARQFWARARVNATPACEERFVKSWRTYLNSVIQQAERRDAKYICTMEEYLHARRDNIGSDPSFALLEITLEVDLPHEVMEHPTIVALARDTTDMIVLANDMCSYKKEIIADDANYNAVTVVMHNHNTNVDGGIQWISDYHDTIVDHFLRLREDVRLKQNGFPSWGARIDREVEAYVEGLGLWIRGHDEWNFGSGRYFGDEGLEVQKSRIVECTVSADPFTPFLIQQEEEVDENAA
CpSTS2	MMQFYLPALVEQCPIEGGTNPHYEQGAAESRAWINGYNVFTDRKRAFFILGSNELLCSHVYYYAGSEEFRTSCDFVNVLFIF**DELSD**EQTGKDALETGQIFLNAMKDERWDDRSKFSSMTKEFRKRFLRRSGPRGTARFLKHWETYCAAVIREAELREMDEVLDLEDFINLRRENSAVRLCYGLIEYCYGIDLPEAVYEDSTFMDIYWAAVDLVCWT**NDVYSYDME**QSKGIAGNNIVTVLMRNRNMSLQQASNYIGQHCETLMDRFVSSQVRLPSWGPVVDREVRLYIQGLGAWIKGNLDWSFETQRYFGPMHEEVKSTRLVTLRPRERIEECDSDSDSDFE	MMQFYLPALVEQCPIEGGTNPHYEQGAAESRAWINGYNVFTDRKRAFFILGSNELLCSHVYYYAGSEEFRTSCDFVNVLFIFDELSDEQTGKDALETGQIFLNAMKDERWDDRSKFSSMTKEFRKRFLRRSGPRGTARFLKHWETYCAAVIREAELREMDEVLDLEDFINLRRENSAVRLCYGLIEYCYGIDLPEAVYEDSTFMDIYWAAVDLVCWTNDVYSYDMEQSKGIAGNNIVTVLMRNRNMSLQQASNYIGQHCETLMDRFVSSQVRLPSWGPVVDREVRLYIQGLGAWIKGNLDWSFETQRYFGPMHEEVKSTRLVTLRPRERIEECDSDSDSDFE	MMQFYLPALVEQCPIEGGTNPHYEQGAAESRAWINGYNVFTDRKRAFFILGSNELLCSHVYYYAGSEEFRTSCDFVNVLFIFDELSDEQTGKDALETGQIFLNAMKDERWDDRSKFSSMTKEFRKRFLRRSGPRGTARFLKHWETYCAAVIREAELREMDEVLDLEDFINLRRENSAVRLCYGLIEYCYGIDLPEAVYEDSTFMDIYWAAVDLVCWTNDVYSYDMEQSKGIAGNNIVTVLMRNRNMSLQQASNYIGQHCETLMDRFVSSQVRLPSWGPVVDREVRLYIQGLGAWIKGNLDWSFETQRYFGPMHEEVKSTRLVTLRPRERIEECDSDSDSDFE
CpSTS3	MVATTASTQPDHFVLPDLVSHCSFPLVYHTDGDRIAAQSVNWLDSNCPDLNAKQRVALRGLQAGELTAFCYNTCTPERLRVVSDFMNYLFHL**DNISD**GMMTRETDVLADVVMNALWFSGKYMPTKEQSADELNPGKLARDFWARCIPDAGPGCQARFKETLELFFEAVNIQARARDDDVVPDLESYIDVRRDTSGCKPCWALIEYALDIDLPDFVVEHPIIEALNQSTNDLVTWS**NDIFSYNVE**QSRGDTHNMIVILMKYHGHTLQSAVDYVGDLCQKTIDDFQANRQKLPSWGAEVDEMVQRYVVGLQDWIVGSLHWSFQTHRYFGADGANVKKNRIVKLLPLKA	MVATTASTQPDHFVLPDLVSHCSFPLVYHTDGDRIAAQSVNWLDSNCPDLNAKQRVALRGLQAGELTAFCYNTCTPERLRVVSDFMNYLFHLDNISDGMMTRETDVLADVVMNALWFSGKYMPTKEQSADELNPGKLARDFWARCIPDAGPGCQARFKETLELFFEAVNIQARARDDDVVPDLESYIDVRRDTSGCKPCWALIEYALDIDLPDFVVEHPIIEALNQSTNDLVTWSNDIFSYNVEQSRGDTHNMIVILMKYHGHTLQSAVDYVGDLCQKTIDDFQANRQKLPSWGAEVDEMVQRYVVGLQDWIVGSLHWSFQTHRYFGADGANVKKNRIVKLLPLKA	MVATTASTQPDHFVLPDLVSHCSFPLVYHTDGDRIAAQSVNWLDSNCPDLNAKQRVALRGLQAGELTAFCYNTCTPERLRVVSDFMNYLFHLDNISDGMMTRETDVLADVVMNALWFSGKYMPTKEQSADELNPGKLARDFWARCIPDAGPGCQARFKETLELFFEAVNIQARARDDDVVPDLESYIDVRRDTSGCKPCWALIEYALDIDLPDFVVEHPIIEALNQSTNDLVTWSNDIFSYNVEQSRGDTHNMIVILMKYHGHTLQSAVDYVGDLCQKTIDDFQANRQKLPSWGAEVDEMVQRYVVGLQDWIVGSLHWSFQTHRYFGADGANVKKNRIVKLLPLKA
CpSTS4	MSESNQHLRIPHTLTAWPWPRTLNPYYQTVKAESSAWLESFKAFDPKAQAGFNSCDFNLLASLAYPLASKEHLRTGCDLMNLFFVI**DEYTD**IEDEVHAAMVASVTMDALRNPFKPRPRGEIIIGEIARQFWARTVPTITEASHRRFIETFDTYLQSVVVQARDRSKQHIRSIQDYLHMRRDNIGAKPSFAILELSLDLPDYVMSHPSIQTATVTAIDMLIIG**NDLCSFRNE**HARGDDTHNILTIARHEFGQGLGHAVNWIESYNKSLRRSFLSAIERVPSWGEEIDTQVAEYLYGLANWVRANDCWSFESHRYFGKHGREIQVHRVVDLACGHPECVQDTDNEVEKPATKGLLYEVEKKTYVGPLQVNTKDPIAFRSDFPLSARPKDITSTPTRLTCVALLCTLSTFIGWLVKL	IPHTLTAWPWPRTLNPYYQTVKAESSAWLESFKAFDPKAQAGFNSCDFNLLASLAYPLASKEHLRTGCDLMNLFFVIDEYTDIEDEVHAAMVASVTMDALRNPFKPRPRGEIIIGEIARQFWARTVPTITEASHRRFIETFDTYLQSVVVQARDRSKQHIRSIQDYLHMRRDNIGAKPSFAILELSLDLPDYVMSHPSIQTATVTAIDMLIIGNDLCSFRNEHARGDDTHNILTIARHEFGQGLGHAVNWIESYNKSLRRSFLSAIERVPSWGEEIDTQVAEYLYGLANWVRANDCWSFESHRYFGKHGREIQVHRVVDLACGHPECVQDTDNEVEKPATKGLLYEVEKKTYVGPLQVNTKDPIAFRSDFPLSARPKDITSTPTRLTCVALLCTLSTFIGWLVKL	MSESNQHLRIPHTLTAWPWPRTLNPYYQTVKAESSAWLESFKAFDPKAQAGFNSCDFNLLASLAYPLASKEHLRTGCDLMNLFFVIDEYTDIEDEVHAAMVASVTMDALRNPFKPRPRGEIIIGEIARQFWARTVPTITEASHRRFIETFDTYLQSVVVQARDRSKQHIRSIQDYLHMRRDNIGAKPSFAILELSLDLPDYVMSHPSIQTATVTAIDMLIIGNDLCSFRNEHARGDDTHNILTIARHEFGQGLGHAVNWIESYNKSLRRSFLSAIERVPSWGEEIDTQVAEYLYGLANWVRANDCWSFESHRYFGKHGREIQVHRVVDLACGHPECVQDTDNEVEKPATKGLLYEVEKKTY**ISHYPLDLRTLPLHQLD**
CpSTS5	MSTIATGVALKSTRRPTKFILPDLVSHCTFKLRCSRFRQKVTTETKQWLFKGGNLGAKDRVKFHGLKAGLLTAMTYPGAAYPQLRVCNDFLTYLFHL**DNLSD**DMDNRGTHTVGNVVMNSLYHPHTYNEPERVGKMARDYFKRVIVTSSTGAQQRFRETMDFFFQSITEQALDRKNGVIPDLESYITLRRDTSGCKPCWALIEYAYNLHIPDEVMEHPLILSLGEAANDLVTWS**NDIFSYNVE**QSKGDTHNMIPVVMNQEGLDLQSAVDFVGQMCKQSIDRFVEDASNLPSWGPEIDRDVAVYVNGLADWIVGSLHWSFESERYFGKTGLEVKATRVVELLPQRA	MSTIATGVALKSTRRPTKFILPDLVSHCTFKLRCSRFRQKVTTETKQWLFKGGNLGAKDRVKFHGLKAGLLTAMTYPGAAYPQLRVCNDFLTYLFHLDNLSDDMDNRGTHTVGNVVMNSLYHPHTYNEPERVGKMARDYFKRVIVTSSTGAQQRFRETMDFFFQSITEQALDRKNGVIPDLESYITLRRDTSGCKPCWALIE**CALLRLLPLFDLTRSFHFLTD**AYNLHIPDEVMEHPLILSLGEAANDLVTWSNDIFSYNVEQSKGDTHNMIPVVMNQEGLDLQSAVDFVGQMCKQSIDRFVEDASNLPSWGPEIDRDVAVYVNGLADWIVGSLHWSFESERYFGKTGLEVKATRVVELLPQRA	MSTIATGVALKSTRRPTKFILPDLVSHCTFKLRCSRFRQKVTTETKQWLFKGGNLGAKDRVKFHGLKAGLLTAMTYPGAAYPQLRVCNDFLTYLFHLDNLSDDMDNRGTHTVGNVVMNSLYHPHTYNEPERVGKMARDYFKRVIVTSSTGAQQRFRETMDFFFQSITEQALDRKNGVIPDLESYITLRRDTSGCKPCWALIEYAYNLHIPDEVMEHPLILSLGEAANDLVTWSNDIFSYNVEQSKGDTHNMIPVVMNQEGLDLQSAVDFVGQMCKQSIDRFVEDASNLPSWGPEIDRDVAVYVNGLADWIVGSLHWSFESERYFGKTGLEVKATRVVELLPQRA
CpSTS6	MLTTPYFTIPDTLRNWPWKRVLNPHQKVCEEEAADWMRSCGAFTPKSQNAFDRCSFGLLGSLAYPRLGRDGLRIACDLMNMFFVI**DEYSD**VASGREARLQADIVMDALYNPLVPRPVGEWIGGEVTRQFWANAIKTATPSSQRRFVRNFQRYVDAVVQQAQDREAHCIRDVKSYFILRRQTIGAIPSLDLLTLEMDLEDEVLDHPIIAKLLELCVDMILIG**NDLYSYNVE**QARGDDTHNFVRIVKDERKCNLNDALRWISDYHDRLADEFLNLMHNLPSFGSKLIDEQVKTYVDGLGNWVRANECWSFESERYFGKKGKLYQKSRRVRLLPSSIALQGQNIDPQAAVEISDLPAAAGIIV	MLTTPYFTIPDTLRNWPWKRVLNPHQKVCEEEAADWMRSCGAFTPKSQNAFDRCSFGLLGSLAYPRLGRDGLRIACDLMNMFFVIDEYSDVASGREARLQADIVMDALYNPLVPRPVGEWIGGEVTRQFWANAIKTATPSSQRRFVRNFQRYVDAVVQQAQDREAHCIRDVKSYFILRRQTIGAIPSLDLLTLEMDLEDEVLDHPIIAKLLELCVDMILIGNDLYSYNVEQARGDDTHNFVRIVKDERKCNLNDALRWISDYHDRLADEFLNLMHNLPSFGSKLIDEQVKTYVDGLGNWVRANECWSFE**VSLSKASEKTWDLTHKQ**SERYFGKKGKLYQKSRRVRLLPSSIALQGQNIDPQAAVEISDLPAAAGIIV	MLTTPYFTIPDTLRNWPWKRVLNPHQKVCEEEAADWMRSCGAFTPKSQNAFDRCSFGLLGSLAYPRLGRDGLRIACDLMNMFFVIDEYSDVASGREARLQADIVMDALYNPLVPRPVGEWIGGEVTRQFWANAIKTATPSSQRRFVRNFQRYVDAVVQQAQDREAHCIRDVKSYFILRRQTIGAIPSLDLLTLEMDLEDEVLDHPIIAKLLELCVDMILIGNDLYSYNVEQARGDDTHNFVRIVKDERKCNLNDALRWISDYHDRLADEFLNLMHNLPSFGSKLIDEQVKTYVDGLGNWVRANECWSFESERYFGKKGKLYQKSRRVRLLPSSIALQGQNIDPQAAVEISDLPAAAGIIV
CpSTS7	MTTSGFRIPNLVVAWPWLRAVNPHMDQTSQNSSSWLENFHAYPPATQENFNRCDF**GLLAALTYPIASK**EHLRTGTDFMNLLFFI**DEITE**IGDETFSTEVLYIAMDALRNPSKLRPKGEHVVGEITRQFWERSIPHITATSHRRFVCALEAYLKSNVVQARDRSSQHIRRIEEYFSTRRDNIGTKPAFVLLELGLDIPDCVMTHPSIQSAVTSLTDMLIIA**NDVFSFKVE**HSRDDDANNVLKVVMCELRCNLARAIEWADERNQQLRTTFLSSIAEVPSWGGQVDIQVSEYLYGLANWVRANECWSFESQRYFGTHGPTVQKDRWVSLDCGHPQCITGGTTTPRTSYRPYLVGMLGIATYALSQYGLMSRATLLRR	MTTSGFRIPNLVVAWPWLRAVNPHMDQTSQNSSSWLENFHAYPPATQENFNRCDFEHLRTGTDFMNLLFFIDEITEIGDETFSTEVLYIAMDALRNPSKLRPKGEHVVGEITRQFWERSIPHITATSHRRFVCALEAYLKSNVVQARDRSSQHIRRIEEYFSTRRDNIGTKPAFVLLELGLDIPDCVMTHPSIQSAVTSLTDMLIIANDVFSFKVEHSRDDDANNVLKVVMCELRCNLARAIEWADERNQQLRTTFLSSIAEVPSWGGQVDIQVSEYLYGLANW**FVFIYEGTGPNGRKPREAEQPLERVMGMDALVFFHIIPWRSPVSETAYWSSVCTGSSIVFDANNLPGGTLDPGED**	No ORF was predicted

Table S5. Continued.

Gene name	Active protein sequence	Predicted sequence by Augustus (*A. oryzae*)	Predicted sequence by Augustus (*C. cinereus*)
CpSTS8	MSNTFIVLPNLEETVYSAFPDHGLNPHYDAVCPHSRAWIKSYSDPIFGPKMRDFMEKCDFELFAAYICPRASPEALRTSMDIGNLLWLY**DEYTD**AEDSVTVKKAEAIVVRTLQDRSFDDGSWICRLVKEFLDLHVRKKAGPNVSRRFVGHFVDYVVRVSDEATQRERHEVLDIEAYVERRRESGAIRLTFDLIESGLHIDLPQYVHEDPAFIAGYNATMDLACWV**NDVYSYNME**QAKGHEASNILTVLMKYEHLDLQAAVEYVAKHCEVLAAQFIEAHANLLARSDPNFSEDAARVLDALGDAVIGNDRWSFETERYFGKDYKAVKQSRIVKLAGRAEGKHALRN	*(MSNTFIVLPNLEETVYSAFPDHGLNPHYDAVCPHSRAWIKSYSDPIFGPK)*MRDFMEKCDFELFAAYICPRASPEALRTSMDIGNLLWLYDEYTDAEDSVTVKKAEAIVVRTLQDRSFDDGSWICRLVKEFLDLHVRKKAGPNVSRRFVGHFVDYVVRVSDEATQRERHEVLDIEAYVERRRESGAIRLTFDLIESGLHIDLPQYVHEDPAFIAGYNATMDLACWVNDVYSYNMEQAKGHEASNILTVLMKYEHLDLQAAVEYVAKHCEVLAAQFIEAHANLLARSDPNFSEDAARVLDALGDAVIGNDRWSFETERYFGKDYKAVKQSRIVKLA**DKPPQQVNFLDVMGIVIKVVLNGPFASIM**	*(MSNTFIVLPNLEETVYSAFPDHGLNPHYDAVCPHSRAWIKSYSDPIFGPK)*MRDFMEKCDFELFAAYICPRASPEALRTSMDIGNLLWLYDEYTDAEDSVTVKKAEAIVVRTLQDRSFDDGSWICRLVKEFLDLHVRKKAGPNVSRRFVGHFVDYVVRVSDEATQRERHEVLDIEAYVERRRESGAIRLTFDLIESGLHIDLPQYVHEDPAFIAGYNATMDLACWVNDVYSYNMEQAKGHEASNILTVLMKYEHLDLQAAVEYVAKHCEVLAAQFIEAHANLLARSDPNFSEDAARVLDALGDAVIGNDRWSFETERYFGKDYKAVKQSRIVKLAGRAEGKHALRN
CpSTS9	MTVLQFVFSVFTSSKPQTFFLPRLDETFKVLPNNGLNPHFSTVRPQSRAWIKQYDGEVCGPKMRAFMDNCNFELSNAYTYPYAEPAGLRAAMDLTNILWLY**DEYTD**TEDGASAERAAVIVNRALKEPGFNDGSWICRMMQDFRRRHIDKAGPDVARRFITNFCNYVDRVAREAILREKNEVLDIPSYILFRRETSAVKTCFDIVEYCLGLNLPQYVHDDPVFVSGYNAAMDLVFWA**NDLFSYNME**QAKGHSGANIVTVIMKSKRVDLQTAADFVAGYCEALTAQLLEAKQILSSRSDPAYSKDAVRVLEAFGDWVRGNDDWSFATERYFGKENNAVRESRIVVVRAPFDETVKLVE	MTVLQFVFSVFTSSKPQTFFLPRLDETFKVLPNNGLNPHFSTVRPQSRAWIKQYDGEVCGPKMRAFMDNCNFELSNAYTYPYAEPAGLRAAMDLTNILWLYDEYTDTEDGASAERAAVIVNRALKEPGFNDGSWICRMMQDFRRRHIDKAGPDVARRFITNFCNYVDRVAREAILREKNEVLDIPSYILFRRETSAVKTCFDIVEYCLGLNLPQYVHDDPVFVSGYNAAMDLVFWANDLFSYNMEQAKGHSGANIVTVIMKSKRVDLQTAADFVAGYCEALTAQLLEAKQILSSRSDPAYSKDAVRVLEAFGDWVRGNDDWSFATERYFGKENNAVRESRIVVVRAPFDETVKLVE	MTVLQFVFSVFTSSKPQTFFLPRLDETFKVLPNNGLNPHFSTVRPQSRAWIKQYDGEVCGPKMRAFMDNCNFELSNAYTYPYAEPAGLRAAMDLTNILWLYDEYTDTEDGASAERAAVIVNRALKEPGFNDGSWICRMMQDFRRRHIDKAGPDVARRFITNFCNYVDRVAREAILREKNEVLDIPSYILFRRETSAVKTCFDIVEYCLGLNLPQYVHDDPVFVSGYNAAMDLVFWANDLFSYNMEQAKGHSGANIVTVIMKSKRVDLQTAADFVAGYCEALTAQLLEAKQILSSRSDPAYSKDAVRVLEAFGDWVRGNDE***(WSFATERYFGKENNAVRESRIVVVRAPFDETVKLVE)***
CpSTS10*	MAPSYQLPDLPSLSRSFELRANPACKITTKASEASLIESKSTNGSHVLSSLERERLAAMKVGLLAAICFPGSDPTQLRLLTDFLTLTVLATRRVKYATLSPQVLTYWTTVDDNDGVWSLVHHDMFNCLKEPLERLASKANQNWKARFGSSVKSFRIAQCSSVTEDMILDPEIETYLSMNRDLSGLSMIFDLFELTQNLTLTVTDETITRTLDKLKVLATDIVSCSVDVAAFNYDQARGNEKNLISLLMRHKRLSVQGALNYAGTLIKQYIDAFMAEERSLLDPTPSPPSNGSSLIPSWIPFTPLVASATAIPPSTPPDPVSKADLTVYVQMLRDCIVGTLNWIYETDLYFGKKGEEIRTFGWIFLSPKVQNNDSEREVPGDNVLP	MAPSYQLPDLPSLSRSFELRANPACKITTKASEASLIESKSTNGSHVLSSLERERLAAMKVGLLAAICFPGSDPTQLRLLTDFLTLTVLATRRVKYATLSPQVLTYWTTVDDNDGVWSLVHHDMFNCLKEPLERLASKANQNWKARFGSSVKSFRIAQCSSVTEDMILDPEIETYLSMNRDLSGLSMIFDLFELTQNLTLTVTDETITRTLDKLKVLATDIVSCSVDVAAFNYDQARGNEKNLISLLMRHKRLSVQGALNYAGTLIKQYIDAFMAEERSLLDPTPSPPSNGSSLIPSWIPFTPLVASATAIPPSTPPDPVSKADLTVYVQMLRDCIVGTLNWIYETDLYFGKKGEEIRTFGWIFLSPKVQNNDSEREVPGDNVLP	MAPSYQLPDLPSLSRSFELRANPACKITTKASEASLIESKSTNGSHVLSSLERERLAAMKVGLLAAICFPGSDPTQLRLLTDFLTLTVLATRRVKYATLSPQVLTYWTTVDDNDGVWSLVHHDMFNCLKEPLERLASKANQNWKARFGSSVKSFRIAQCSSVTEDMILDPEIETYLSMNRDLSGLSMIFDLFELTQNLTLTVTDETITRTLDKLKVLATDIVSCSVDVAAFNYDQARGNEKNLISLLMRHKRLSVQGALNYAGTLIKQYIDAFMAEERSLLDPTPSPPSNGSSLIPSWIPFTPLVASATAIPPSTPPDPVSKADLTVYVQMLRDCIVGTLNWIYETDLYFGKKGEEIRTFGWIFLSPKVQNNDSEREVPGDNVLP
CpSTS11	MSQKSPSTFRIPDLEAIFSAFPDEGTSPYYDDVLPEARAWIKLYQDQVYGPKMTEFIDRCKIELITYYVHPVASRSCVRAMMDLHNLFWLY**DEATD**VQSGQTAQETAKVVRNSLTSPEFNDGSWLCEMLQDFRKRHLDGVRSSSFVARFIEHFCFYTDRVADEAIYREKQRVLDIPSYMAFRRETAAVRVVMDTVEYCAELELPRTVLDDPVFQVAYDAALDLAFGT**NDIHSYNME**QSKDHSGANVITVIMKERGLDVQGAMDYFGGYCEALTAQFLDAKRKIEKREGQEWKDAVLILDGYTHFLTGQVRWGFATERYFGKKNKEVQETRVVELRAPFVDHVDLAD	MSQKSPSTFRIPDLEAIFSAFPDEGTSPYYDDVLPEARAWIKLYQDQVYGPKMTEFIDRCKIELITYYVHPVASRSCVRAMMDLHNLFWLYDEATDVQSGQTAQETAKVVRNSLTSPEFNDGSWLCEMLQDFRKRHLDGVRSSSFVARFIEHFCFYTDRVADEAIYREKQRVLDIPSYMAFRRETAAVRVVMDTVEYCAELELPRTVLDDPVFQVAYDAALDLAFGTNDIHSYNMEQSKDHSGANVITVIMKERGLDVQGAMDYFGGYCEALTAQFLDAKRKIEKREGQEWKDAVLILDGYTHFLTGQVRWGFATERYFGKKNKEVQETRVVELRAPFVDHVDLAD	MSQKSPSTFRIPDLEAIFSAFPDEGTSPYYDDVLPEARAWIKLYQDQVYGPKMTEFIDRCKIELITYYVHPVASRSCVRAMMDLHNLFWLYDEATDVQSGQTAQETAKVVRNSLTSPEFNDGSWLCEMLQDFRKRHLDGVRSSSFVARFIEHFCFYTDRVADEAIYREKQRVLDIPSYMAFRRETAAVRVVMDTVEYCAELELPRTVLDDPVFQVAYDAALDLAFGTNDIHSYNMEQSKDHSGANVITVIMKERGLDVQGAMDYFGGYCEALTAQFLDAKRKIEKREGQEWKDAVLILDGYTHFLTGQVRWGFATERYFGKKNKEVQETRVVELRAPFVDHVDLAD
CpSTS12	MESLPVTLVLPRLDDIFDDLPNNVVNPNYSIACPASRLWIEQYGTQIYGPKMQAFMNNCNFELSTTYTYPYADAKRLRATMDLVNILWLY**DEYTD**RASGAHAKEMASIVYQALSGQQVAANSWVGCMMQDFYRQHIEKAGPNTSRRFVDHFCRYAQQVGEEASLREQRQILNMHEYIDFRRETSGVRSCFDLVEYCLGIDLPQFVHDDPIFTMGYNAAMDLVFWV**NDLYSYNME**QAKGHGAANVVTVIMKSKSLGLQAAVDHLADACEVLTAQFLEAKSRLSKHPESIFSKEAVKVLDAYGDWVRGNEEWSFVTERYFGKENKAVRQSRMVKIKTPFGEMAPFSGRKNYL	MESLPVTLVLPRLDDIFDDLPNNVVNPNYSIACPASRLWIEQYGTQIYGPKMQAFMNNCNFELSTTYTYPYADAKRLRATMDLVNILWLYDEYTDRASGAHAKEMASIVYQALSGQQVAANSWVGCMMQDFYRQHIEKAGPNTSRRFVDHFCRYAQQVGEEASLREQRQILNMHEYIDFRRETSGVRSCFDLVEYCLGIDLPQFVHDDPIFTMGYNAAMDLVFWVN**VSLEKGSNSSVAYRPSQ**DLYSYNMEQAKGHGAANVVTVIMKSKSLGLQAAVDHLADACEVLTAQFLEAKSRLSKHPESIFSKEAVKVLDAYGDWVRGNEEWSFVTERYFGKENKAVRQSRMVKIKTPFGEMAPFSGRKNYL	MESLPVTLVLPRLDDIFDDLPNNVVNPNYSIACPASRLWIEQYGTQIYGPKMQAFMNNCNFELSTTYTYPYADAKRLRATMDLVNILWLYDEYTDRASGAHAKEMASIVYQALSGQQVAANSWVGCMMQDFYRQHIEKAGPNTSRRFVDHFCRYAQQVGEEASLREQRQILNMHEYIDFRRETSGVRSCFDLVEYCLGIDLPQFVHDDPIFTMGYNAAMDLVFWVNDLYSYNMEQAKGHGAANVVTVIMKSKSLGLQAAVDHLADACEVLTAQFLEAKSRLSKHPESIFSKEAVKVLDAYGDWVRGNEEWSFVTERYFGKENKAVRQSRMVKIKTPFGEMAPFSGRKNYL
CpSTS13	MPQDPPTTFRIPDLETIFSVFPDEGTNPHYDDVFPEAREWILRYSKHVYGPKMAEFIDSCKVELLTCYIYPSASKSRLRSMMDLYNLFWIYDETTDVQTGQEAKETAQVVRNALTNPDFSDDSWLCAILQDFRRRNLDGIMSPGFVLRFIDHFCDYANGVADEALFREKDLMLDIQEYLKFRREAAAVRVVLDSVEYCLDLELEQNVMDDPVFQMAYNAALDLAFGT**NDIQSYNME**QAKQHKGANIISVIMRARTLDLQGAMDYFGGYCQALTAQFLDAKCIIEKRVDRPEWKDAVRILEGYGCFLAGQVRWGFTTERYFGKKNKEVESTKVVELRAPFVDHVHISD	*(MPQDPPTTFRIPDLETIFSVFPDEGTNPHYDDVFPEAREWILRYSKHVYGPKMAEFIDSCKVELLTCYIYPSASKSRLRSMMDLYNLFWIYDETTDVQTGQEAKETAQVVRNALTNPDFSDDSWLCAILQDFRRRNLDGI)*MSPGFVLRFIDHFCDYANGVADEALFREKDLMLDIQEYLKFRREAAAVRVVLDSVEYCLDLELEQNVMDDPVFQMAYNAALDLAFGTNDIQSYNMEQAKQHKGANIISVIMRARTLDLQGAMDYFGGYCQALTAQFLDAKCIIEKRVDRPEWKDAVRILEGYGCFLAGQ**LTECESLTQGGIAGFFYELQALALQSRSIFSPGCIRDDDTFKSLVSKTDEFSGTSLARFTCTYHMPEMAYMACHYHAYDAEFRVDPTSEANGCIAPTTRCEEIPLS**	MPQDPPTTFRIPDLETIFSVFPDEGTNPHYDDVFPEAREWILRCKVELLTCYIYPSASKSRLRSMMDLYNLFWIYDETTDVQTGQEAKETAQVVRNALTNPDFSDDSWLCAILQDFRRRNLDGIMSPGFVLRFIDHFCDYANGVADEALFREKDLMLDIQEYLKFRREAAAVRVVLDSVEYCLDLELEQNVMDDPVFQMAYNAALDLAFGTNDIQSYNMEQAKQHKGANIISVIMRARTLDLQGAMDYFGGYCQALTAQFLDAKCIIEKRVDRPEWKDAVRILEGYGCFLAGQVRQVWGFTTERYFGKKNKEVESTKVVELRAP**LLLSSLVASLVLDAFAAEYCCPTDDDTFKSLVSKTDEFSGTSLARFTCTYHMPEMAYMACHYHAYDAEFRVDPTSEANGCIAPTTRCEEIPLS**
CpSTS14	MNVMNTPITTATMPVSPPCPNPDPTFQTRLSARWHPSTAEIVPKVRQYIVQNWPWSSEESKRRYLTGNAEDGVMYAFPGVKNDRIEAVTTWLALHFLI**DDYILE**SVSDSAASGLDPGKKKEAIQRLFGIMRRLVRPNVNNPVEVMVDHVASSFLSCSSEDEVAQRHARQILESTIQFITAAENGDGKATAMGDLNSYLAYRMVEAGVFLSLDLAFWGGEIYIPSHISNDPQIRSFFKLVSDHLVLV**NDIYSYKVE**QDRSEGVGGAFNAVSVLMRSKHVDAQDAMGIIKTQLTTLEEATLEEAAVLRDRYAEKGEEEEVVDKMIQTILEMMAGNCEWSKFCGRYNSSPVSDEASVGTDRALEQGTTSLS	*(MNVMNTPITTAT)*MPVSPPCPNPDPTFQTRLSARWHPSTAEIVPKVRQYIVQNWPWSSEESKRRYLTGNAEDGVMYAFPGVKNDRIEAVTTWLALHFLIDDYILESVSDSAASGLDPGKKKEAIQRLFGIMRRLVRPNVNNPVEVMVDHVASSFLSCSSEDEVAQRHARQILESTIQFITAAENGDGKATAMGDLNSYLAYRMVEAGVFLSLDLAFWGGEIYIPSHISNDPQIRSFFKLVSDHLVLVNDIYSYKVEQDRSEGVGGAFNAVSVLMRSKHVDAQDAMGIIKTQLTTLEEATLEEAAVLRDRYAEKGEEEEVVDKMIQTILEMMAGNCEWSKFCGRYNSSPVSDEASVGTDRALEQGTTSLS	MNVMNTPITTATMPVSPPCPNPDPTFQTRLSARWHPSTAEIVPKVRQYIVQNWPWSSEESKRRYLTGNAEDGVMYAFPGVKNDRIEAVTTWLALHFLIDDYILESVSDSAASGLDPGKKKEAIQRLFGIMRRLVRPNVNNPVEVMVDHVASSFLSCSSEDEVAQRHARQILESTIQFITAAENGDGKATAMGDLNSYLAYRMVEAGVFLSLDLAFWGGEIYIPSHISNDPQIRSFFKLVSDHLVLVNDIYSYKVEQDRSEGVGGAFNAVSVLMRSKHVDAQDAMGIIKTQLTTLEEATLEEAAVLRDRYAEKGEEEEVVDKMIQTILEMMAGNCEWSKFCGRYNSSPVSDEASVGTDRALEQGTTSLS

* Amino acid sequence based on the recovered cDNA sequence from AO-transformant is shown in the left column, demonstrating that CpSTS10 lacks the catalytic motifs (DDxxD and NSE motifs).

Table S5. Continued.

Gene name	Active protein sequence	Predicted sequence by Augustus (*A. oryzae*)	Predicted sequence by Augustus (*C. cinereus*)
CpSTS15*	MSSHRPTQQPNPLVDPPDFVSKFLAHDREDRDEIITSVQAYFLAHWDWPSPTRKAWYSKADLEDWTTLMCPAGPSERAWAFTAYVTFWFLY**DDLME**IMGPEELGSSIPRIARILHGEEGLEDMTTPEFILKEICLKINTIAPGKRIFEATLVYMKAATAKQERLQSLTLGFDAYLKYRMLDAACWLTLEGAYWVQEITIPEHLKTEKIWLLQELSLFHGILI**NDLFSYRKE**VKASSQEVQDHDKTLYNGLVILMQSYQFTLLEALEEMKKKIWDYEDEFMAVLDDVRETHKSSAEDREFIERLAVALMDVIGGNVAWSACCGRYNRL	MSSHRPTQQPNPLVDPPDFVSKFLAHDREDRDEIITSVQAYFLAHWDWPSPTRKAWYSKADLEDWTTLMCPAGPSERAWAFTAYVTFWFLYDDLMEIMGPEELGSSIPRIARILHGEEGLEDMTTPEFILKEICLKINTIAPGKRIFEATLVYMKAATAKQERLQSLTLGFDAYLKYRMLDAACWLTLEGAYWVQEITIPEHLKTEKIWLLQELSLFHGILINDLFSYRKEVKASSQEVQDHDKTLYNGLVILMQSYQFTLLEALEEMKKKIWDYEDEFMAVLDD**RSFLPPLNTKLKMELADVHALLNLNCLSGLSLITLGTEYLVFRRPPPTSPRSFPRTPPPDHSLLSHYQGNSRLHEAALTYGYDHLVPILVEVLVFGPLTAQLSGNGYGNNEKVNDRFKFRVASDIKVIQLLA**	MSSHRPTQQPNPLVDPPDFVSKFLAHDREDRDEIITSVQAYFLAHWDWPSPTRKAWYSKADLEDWTTLMCPAGPSERAWAFTAYVTFWFLYDDLMEIMGPEELGSSIPRIARILHGEEGLEDMTTPEFILKEICLKINTIAPGKRIFEATLVYMKAATAKQERLQSLTLGFDAYLKYRMLDAACWLTLEGAYWVQEITIPEHLKTEKIWLLQELSLFHGILINDLFSYRKEVKASSQEVQDHDKTLYNGLVILMQSYQFTLLEALEEMKKKIWDYEDEFMAVLDD**RSFLPPLNTKLKMELADVHALLNLNCLLAPGPTINTWFFVDLHRPHLVPFPELLLQTTHSYPTTKATVAYTRQL**
CpSTS16	MPLPVTFRLPDLSSIFSIFPDLGINSHYECTYPESREWIAQYHTGVYGPKMRDFMERGKIELLGAYTYPYASKERLRFVMDLHNVSWLF**DESSD**VKTGQQAGVTAVVFRRALIDPEFNDGSWLCHMLKDFRQRHLDNVMSHAFVEMFIRDFIGYADGMSAEACYRDLNKVLDIAGYVKLRREAGAVRLALHSVEYCLEKELPSYVREDPAFVIAYNGALDLGYIV**NDIHSYNME**QSKGHGHHAANIVTVFMEAQNIGLQAAMDYAGGFCHGLVLQILEAKELLAARSDPVFSNDAVKVIEGCINFFKGQDIWDFESERFFGKSKDAVRKTKVVNLRALFEDSVNLNE	**MVKLGDIIQHIPPSD**MPLPVTFRLPDLSSIFSIFPDLGINSHYECTYPESREWIAQYHTGVYGPKMRDFMERGKIELLGAYTYPYASKERLRFVMDLHNVSWLFDESSDVKTGQQAGVTAVVFRRALIDPEFNDGSWLCHMLKDFRQRHLDNVMSHAFVEMFIRDFIGYADGMSAEACYRDLNKVLDIAGYVKLRREAGAVRLALHSVEYCLEKELPSYVREDPAFVIAYNGALDLGYIVNDIHSYNMEQSKGHGHHAANIVTVFMEAQNIGLQAAMDYAGGFCHGLVLQILEAKELLAARSDPVFSNDAVKVIEGCINFFKGQDIWDFESERFFGKSKDAVRKTKVVNLRALFEDSVNLNE	MPLPVTFRLPDLSSIFSIFPDLGINSHYECTYPESREWIAQYHTGVYGPKMRDFMERGKIELLGAYTYPYASKERLRFVMDLHNVSWLFDESSDVKTGQQAGVTAVVFRRALIDPEFNDGSWLCHMLKDFRQRHLDNVMSHAFVEMFIRDFIGYADGMSAEACYRDLNKVLDIAG*(YVKLRREAGAVRLALHS)*VEYCLEKELPSYVREDPAFVIAYNGALDLGYIVNDIHSYNMEQSKGHGHHAANIVTVFMEAQNIGLQAAMDYAGGFCHGLVLQILEAKELLAARSDPVFSNDAVKVIEGCINFFKGQDIWDFESERFFGKSKDAVRKTKVVNLRALFEDSVNLNE
CpSTS17	MNSDLQSTLASLDTTYTCRFRARYHAKSEQVMTVVQDYFVKRWPWRSEQAKQLYIKTNLEEATCICFPTTLDDRIETVVTWFCYMHIL**DDIIE**GLEANKATKLVDRLSNIILGTLDPDPEDRLGVMAADVCSRFRLGENDGEEKHGLDIISESCKLLRHTSSADSKLQSLSTYDTFIEWRELDVGLWFSTSLFLWGCGIYMRLHMSNDPEVRQLLRVSGRHIAVA**NDLFSYRVE**ALRAGQTSQILLNTIAIIAKEKQVDPQTAMNMTKQRLSEMEEEVEVLIGKLRDRYTGEEGELIERLFVVCKGMMAGNCEWSSICFRYNGAQRAVAA	MNSDLQSTLASLDTTYTCRFRARYHAKSEQVMTVVQDYFVKRWPWRSEQAKQLYIKTNLEEATCICFPTTLDDRIETVVTWFCYMHILDDIIEGLEANKATKLVDRLSNIILGTLDPDPEDRLGVMAADVCSRFRLGENDGEEKHGLDIISESCKLLRHTSSADSKLQSLSTYDTFIEWRELDVGLWFSTSLFLWGCGIYMRLHMSNDPEVRQLLRVSGRHIAVANDLFSYRVEALRAGQTSQILLNTIAIIAKEKQVDPQTAMNMTKQRLSEMEEEVEVLIGKLRDRYTGEEGELIERLFVVCKGMMAGNCEWSSICFRYNGAQRAVAA	MNSDLQSTLASLDTTYTCRFRARYHAKSEQVMTVVQDYFVKRWPWRSEQAKQLYIKTNLEEATCICFPTTLDDRIETVVTWFCYMHILDDIIEGLEANKATKLVDRLSNIILGTLDPDPEDRLGVMAADVCSRFRLGENDGEEKHGLDIISESCKLLRHTSSADSKLQSLSTYDTFIEWRELDVGLWFSTSLFLWGCGIYMRLHMSNDPEVRQLLRVSGRHIAVANDLFSYRVEALRAGQTSQILLNTIAIIAKEKQVDPQTAMNMTKQRLSEMEEEVEVLIGKLRDRYTGEEGELIERLFVVCKGMMAGNCEWSSICFRYNGAQRAVAA
CpSTS18	MVHWNHPPTFVLRNICDITGRVFELKENPRIAEANSAVLKWFEQFNVYDKKKADKFLNVGKFDIFAALSFPEADLEHLTTCLIFFLWAFAT**DDLSD**EGEFQSQPDQVQHGHDISCSILDDDDAPQPDYPYAAMLWDLLRRLRSTGHMGMYKRFKQAFLDFSSSQVQQSTNRNVDRIPPVDEFILMRRKTIGAALVEAMVEYSLDLDIPSYVWEHPVIVGMSQATSDIMTWP**NDLCSFNKE**QADGDYQNLVCVLQHNHGLELQEAIDLLTKMISDRVQDYVDLKNQLPSFGPDVDPALHTYLTALEQFVQGTVVWYYSSPRYFRHLDPRGKPEVLIHLFPKTDAPTLPVVVQEKSRPFIEREIYLPAKRLLGVFVNYVAITVFGYSVYRLYGDS	MVHWNHPPTFVLRNICDITGRVFELKENPRIAEANSAVLKWFEQFNVYDKKKADKFLNVGKFDIFAALSFPEADLEHLTTCLIFFLWAFATDDLSDEGEFQSQPDQVQHGHDISCSILDDDDAPQPDYPYAAMLWDLLRRLRSTGHMGMYKRFKQAFLDFSSSQVQQSTNRNVDRIPPVDEFILMRRKTIGAALVEAMVEYSLDLDIPSYVWEHPVIVGMSQATSDIMTWPNDLCSFNKEQADGDYQNLVCVLQHNHGLELQEAIDLLTKMISDRVQDYVDLKNQLPSFGPDVDPALHTYLTALEQFVQGTVVWYYSSPRYFRHLDPRGKPEVLIHLFPKTDAPTLPVVVQEKSRPFIEREIYLPAKRLLGVFVNYVAITVFGYSVYRLYGDS	MVHWNHPPTFVLRNICDITGRVFELKENPRIAEANSAVLKWFEQFNVYDKKKADKFLNVGKFDIFAALSFPEADLEHLTTCLIFFLWAFATDDLSDEGEFQSQPDQVQHGHDISCSILDDDDAPQPDYPYAAMLWDLLRRLRSTGHMGMYKRFKQAFLDFSSSQVQQSTNRNVDRIPPVDEFILMRRKTIGAALVEAMVEYSLDLDIPSYVWEHPVIVGMSQATSDIMTWPNDLCSFN**AVQ**KEQADGDYQNLVCVLQHNHGLELQEAIDLLTKMISDRVQDYVDLKNQLPSFGPDVDPALHTYLTALEQFVQGTVVWYYSSPRYFRHLDPRGKPEVLIHLFPKTDAPTLPVVVQEKSRPFIEREIYLPAKRLLGVFVNYVAITVFGYSVYRLYGDS
ShSTS1	MESVREHIPRLQHFLGEIGYRHTTPPAPTLDFLHAHHHWIHHVLGPMTSWTVAKLNALEDSSSTIFERAYPLSDAEMKFVLAKLTAIAIFL**DDSLD**DEETYDDIGNFAHRVYLGEAQPTGVLTLYHQGIQELSKMHEGDAVFRGLAVAPWITFIDACMLEKRLLTFDSKLRVSPRDLGYQRLRNGTDFTSLRAPKATPSEVEVSFPIFLRHKSGIGEAYAAAIFKSSRYQELPLSRFVKSMPDMIYYIELV**NDLMSFYKE**QLAGETANLIHLQHQSWKGGQGTGPYGSWTLLDTFSRLCDDTRDAAFRVDELLRLDECEKIANGELRGEEVGLSPMDVTMAAQWREFRDGYVSWHLECQRYKLDFIKLSTFE	MESVREHIPRLQHFLGEIGYRHTTPPAPTLDFLHAHHHWIHHVLGPMTSWTVAKLNALEDSSSTIFERAYPLSDAEMKFVLAKLTAIAIFLDDSLEDEETYDDIGNFAHRVYLGEAQPTGVLTLYHQGIQELSKMHEGDAVFRGLAVAPWITFIDACMLEKRLLTFDSKLRVSPRDLGYQRLRNSTDFTSLRAPKATPSEVEVSFPIFLRHKSGIGEAYAAAIFKSSRYQELPLSRFVKSMPDMIYYIELVNDLMSFYKEQLAGETANLIHLQHQSWKGGQGTGPYGSWTLLDTFSRLCDETRDAAFRVDELLRLDECEKIANGELRGEEVGLSPMDVTMAAQWREFRDGYVSWHLECQRYKLDFIKLSTFE	MESVREHIPRLQHFLGEIGYRHTTPPAPTLDFLHAHHHWIHHVLGPMTSWTVAKLNALEDSSSTIFERAYPLSDAEMKFVLAKLTAIAIFLDDSLEDEETYDDIGNFAHRVYLGEAQPTGVLTLYHQGIQELSKMHEGDAVFRGLAVAPWITFIDACMLEKRLLTFDSKLRVSPRDLGYQRLRNSTDFTSLRAPKATPSEVEVSFPIFLRHKSGIGEAYAAAIFKSSRYQELPLSRFVKSMPDMIYYIELVNDLMSFYKEQLAGETANLIHLQHQSWKGGQGTGPYGSWTLLDTFSRLCDETRDAAFRVDELLRLDECEKIANGELRGEEVGLSPMDVTMAAQWREFRDGYVSWHLECQRYKLDFIKLSTFE
ShSTS3	MSTAKPEIPPMESFPPINVYPREAEIIKYCNNYVAEHFPFNNDAEVKHFNGMEIPAYACRVVSFARDHEKMRKVSILIVFYFIF**DDWVD**KNGMKLDKSTVMALLPPPSEVPIKPPQAGKMTLSDISAELYGAVRDDMPKADYDRMVNDMMEYLRVQRMAPGYNTLQELLDLRSGEVGVYVLFRYIIYAMELSVSGKELDDPLVKRAQVLGSEAGVLR**NEASSYVKE**VNEGSGAHNVITKLQEWSGCTEKEAMKQVLDAIEKRQEELREMCLKVTEAPHLSEDCKTFVKTIPYIVAGNTWWHHHSTRYAEGREVTVP	MSTAKPEIPPMESFPPINVYPREAEIIKYCNNYVAEHFPFNNDAEVKHFNGMEIPAYACRVVSFARDHEKMRKVSILIVFYFIFDDWVDKNGMKLDKSTVMALLPPPSEVPIKPPQAGKMTLSDISAELYGAVRDDMPKADYDRMVNDMMEYLRVQRMAPGYNTLQELLDLRSGEVGVYVLFRYIIYAMELSVSGKELDDPLVKRAQVLGSEAGVLRNEASSYVKEVNEGSGAHNVITKLQEWSGCTEKEAMKQVLDAIEKRQEELREMCLKVTEAPHLSEDCKTFVKTIPYIVAGNTWWHHH**VCLNPSTPALLLADALPLAGKVYEICRGSRGDCALEYGRVAMSLSILF**	MSTAKPEIPPMESFPPINVYPREAEIIKYCNNYVAEHFPFNNDAEVKHFNGMEIPAYACRVVSFARDHEKMRKVSILIVFYFIFDDWVDKNGMKLDKSTVMALLPPPSEVPIKPPQAGKMTLSDISAELYGAVRDDMPKADYDRMVNDMMEYLRVQRMAPGYNTLQELLDLRSGEVGVYVLFRYIIYAMELSVSGKELDDPLVKRAQVLGSEAGVLRNEASSYVKEVNEGSGAHNVITKLQEWSGCTEKEAMKQVLDAIEKRQEELREMCLKVTEAPHLSEDCKTFVKTIPYIVAGNTWWHHHSTRYAEGREVTVP

* Amino acid sequence based on the recovered cDNA sequence from AO-transformant is shown in the left column although it is not active in AO.

Table S5. Continued.

Gene name	Active protein sequence	Predicted sequence by Augustus (*A. oryzae*)	Predicted sequence by Augustus (*C. cinereus*)
ShSTS4	MSETKVGKVAPFPPLPGAPYPPVRNHPRWKELYRLHDEWMMKYWPFSSEKKRARIPFMNLAGFSTWCAPAADFDRMVWGARIAGIFFLA**DDYID**SGKMLDRIPGFKAAATGTGPLHKEDQAEICHDIVFRAIKATSHPRTFDQLTKCTHEWWDSNIHEPFQNLDQYLAVRRVNIAMYFANAYFRYTLDINLTDEQVNHPLMREAEGIVSDHVGLT**NDFFSYLKE**KMTNSDDTNIIRILMDHEHLSYEEAKTVIEKKIRQKEQDFIGAGMAVLNDPELGKDREIYRWIANLQYCMGGNLAWSQESGRYNVGVIDGISFPSLSYAAEPTPEDEVVDDTEESRLRELIFNVKDIPPPDFTIDDDAIFMTNPHSHLQDNVVPLPRPENVGIIGLEVYFPKRCISIDALEDFDGVAKGKYTIGLGQQYLAFTDDREDINSFALSAVSSLLEKYNIDPRSIGRLDVGTETLIDKSKSVKTTLMDLFAASGNHDVEGIDSKNACYGSTAAVLNAVNWIESSSWDGRYAIVFAGDIAIYAEGPARPVGGAGAVALLIGPDAPLVLEPTHGSYMANTYDFYKPRMESEYPVVDGPSSVTTYITALDESFKAYQRKVQEGSSRDVPPPYANGANGKASATKSVKLSDFDYSVFHSPYGKLVQKAYGRLTYHDFVAHPTAPVYRDLPTDILSKDASATLTDKSVEKTFAAASASMYKQVVTPSLLISNRCGNMYTGSLYGGLASLLTSIPSYELFDKRISMFAYGSGCASTFFAIKVRGDTSHIKAKLDLERRLAEMDVRPCEDYVVALKLREETHNAPSYIPNDSDAGLWPGSYRLEAVDGKYRRSYTVTH	MSETKVGKVAPFPPLPGAPYPPVRNHPRWKELYRLHDEWMMKYWPFSSEKKRARIPFMNLAGFSTWCAPAADFDRMVWGARIAGIFFLADDYIDSGKMLDRIPGFKAAATGTGPLHKEDQAEICHDIVFRAIKATSHPRTFDQLTKCTHEWWDSNIHEPFQNLDQYLAVRRVNIAMYFANAYFRYTLDINLTDEQVNHPLMREAEGIV**SGKSLPPISIDAVVIKYHL**SDHVGLTNDFFSYLKEKMTNSDDTNIIRILMDHEHLSYEEAKTVIEKKIRQKEQDFIGAGMAVLNDPELGKDREIYRWIANLQYCMGGNLAWSQESGRYNVGVIDGISFPSLSYAAEPTPEDEVVDDTEESRLRELIFNVKDIPPPDFTIDDDAIFMTNPHSHLQDNVVPLPRPENVGIIGLEVYFPKRCISIDALEDFDGVAKGKYTIGLGQQYLAFTDDREDINSFALSAVSSLLEKYNIDPRSIGRLDVGTETLIDKSKSVKTTLMDLFAASGNHDVEGIDSKNACYGSTAAVLNAVNWIESSSWDGRYAIVFAGDIAIYAEGPARPVGGAGAVALLIGPDAPLVLEPTHGSYMANTYDFYKPRMESEYPVVDGPSSVTTYITALDESFKAYQRKVQEGSSRDVPPPYANGANGKASATKSVKLSDFDYSVFHSPYGKLVQKAYGRLTYHDFVAHPTAPVYRDLPTDILSKDASATLTDKSVEKTFAAASASMYKQVVTPSLLISNRCGNMYTGSLYGGLASLLTSIPSYELFDKRISMFAYGSGCASTFFAIKVRGDTSHIKAKLDLERRLAEMDVRPCEDYVVALKLREETHNAPSYIPNDSDAGLWPGSYRLEAVDGKYRRSYTVTH	MSETKVGKVAPFPPLPGAPYPPVRNHPRWKELYRLHDEWMMKYWPFSSEKKRARIPFMNLAGFSTWCAPAADFDRMVWGARIAGIFFLADDYIDSGKMLDRIPGFKAAATGTGPLHKEDQAEICHDIVFRAIKATSHPRTFDQLTKCTHEWWDSNIHEPFQNLDQYLAVRRVNIAMYFANAYFRYTLDINLTDEQVNHPLMREAEGIVSDHVGLTNDFFSYLKEKMTNSDDTNIIRILMDHEHLSYEEAKTVIEKKIRQKEQDFIGAGMAVLNDPELGKDREIYRWIANLQYCMGGNLAWSQESGRYNVGVIDGISFPSLSYAAEPTPEDEVVDDTEESRLRELIFNVKDIPPPDFTIDDDAIFMTNPHSHLQDNVVPLPRPENVGIIGLEVYFPKRCISIDALEDFDGVAKGKYTIGLGQQYLAFTDDREDINSFALSAVSSLLEKYNIDPRSIGRLDVGTETLIDKSKSVKTTLMDLFAASGNHDVEGIDSKNACYGSTAAVLNAVNWIESSSWDGRYAIVFAGDIAIYAEGPARPVGGAGAVALLIGPDAPLVLEPTHGSYMANTYDFYKPRMESEYPVVDGPSSVTTYITALDESFKAYQRKVQEGSSRDVPPPYANGANGKASATKSVKLSDFDYSVFHSPYGKLVQKAYGRLTYHDFVAHPTAPVYRDLPTDILSKDASATLTDKSVEKTFAAASASMYKQVVTPSLLISNRCGNMYTGSLYGGLASLLTSIPSYELFDKRISMFAYGSGCASTFFAIKVRGDTSHIKAKLDLERRLAEMDVRPCEDYVVALKLREETHNAPSYIPNDSDAGLWPGSYRLEAVDGKYRRSYTVTH
ShSTS5	MELSSLRPFPASFVLPNLANITRQAFNLKLNPHSQSANSAMKSWFKSFHVYDEQKSREFLEAGKFDLYAALSFPDADLQHLETCLAFFFWAFST**DDLSD**EGDLQSKPEEVQVGVDISTSALSSHAPTSLDFPYAAMLQSLFNRIKTTATKGASERFIQAFKDWSSSQVMQSRNRSKLLLPSVEDFILMRRNTIGAALVEAMIEYSLDLDLPDYVFRDPVVIAMSEATTDIMTWP**NDLCSFNKE**QADGDYQNLVCCLMAQYDLGLQDAVDRLVGMISTRVRDYITLKEQLPLFGAEVDTMLRKYHAALEHYVQGTIVWYYSSPRYFHGEQIIEKESTRIILFSKASSKDCS	MELSSLRPFPASFVLPNLANITRQAFNLKLNPHSQSANSAMKSWFKSFHVYDEQKSREFLEAGKFDLYAALSFPDADLQHLETCLAFFFWAFSTDDLSDEGDLQSKPEEVQVGVDISTSALSSHAPTSLDFPYAAMLQ**RFVLYVVYVICTT**FIQAFKDWSSSQVMQSRNRSKLLLPSVEDFILMRRNTIGAALVE**GSQSFASSFSINSEIQIL**AMIEYSLDLDLPDYVFRDPVVIAMSEATTDIMTWPNKEQADGDYQNLVCCLMAQYDLGLQDAVDRLVGMISTRVRDYITLKEQLPLFGAEVDTMLRKYHAALEHYVQGTIVWYYSSPRYFHGEQIIEKESTRIILFSKASSKDCS	MELSSLRPFPASFVLPNLANITRQAFNLKLNPHSQSANSAMKSWFKSFHVYDEQKSREFLEAGKFDLYAALSFPDADLQHLETCLAFFFWAFSTDDLSDEGDLQSKPEEVQVGVDISTSALSSHAPTSLDFPYAAMLQSLFNRIKTTATKGASERFIQAFKDWSSSQVMQSRNRSKLLLPSVEDFILMRRNTIGAALVE**GSQSFASSFSINSEIQIL**AMIEYSLDLDLPDYVFRDPVVIAMSEATTDIMTWPNDLCSFNKEQADGDYQNLVCCLMAQYDLGLQDAVDRLVGMISTRVRDYITLKEQLPLFGAEVDTMLRKYHAALEHYVQGTIVWYYSSPRYFHGEQIIEKESTRIILFSKASSKDCS
ShSTS7	MPHSTVHSHTLISDDSVLLFPDLISYCAYPLRVNPYGRSVADDSERWLLNGAHLSDKKRKAFLRLRAGDLASMCYPDASAKSLRVVADYMNYLFKL**DDWTD**EFEAEDVDGMRDCVLAALRDPLHYETDKAVGKLAKSFFGRFVQHGGPLRTKRFIDTMVLFFRAVRQQALDRTYDDIPDLESYIALRRDTSGCKPCFALIEFAGGYDLPDAVVEHPSIQILQDATNDLVTWS**NDIFSYNVE**QSRGDTHNMVVVLMHEQGLAVQEAIDAVADLCERSIDTFEQTRRSLPSWGPIVDSNVESYIDGLQNWIIGSLHWSFLTERYFGKDRRDVKRKLFVKLLAKRC	MPHSTVHSHTLISDDSVLLFPDLISYCAYPLRVNPYGRSVADDSERWLLNGAHLSDKKRKAFLRLRAGDLASMCYPDASAKSLRVVADYMNYLFKLDDWTDEFEAEDVDGMRDCVLAALRDPLHYETDKAVGKLAKSFFGRFVQHGGPLRTKRFIDTMVLFFRAVRQQALDRTYDDIPDLESYIALRRDTSGCKPCFALIEFAGGYDLPDAVVEHPSIQILQDATNDLVTWSN**VSNPMPSYTRHPLQVCKSRSIADPLRVQ**DIFSYNVEQSRGDTHNMVVVLMHEQGLAVQEAIDAVADLCERSIDTFEQTRRSLPSWGPIVDSNVESYIDGLQNWIIGSLHWSFLTERYFGKDRRDVKRKLFVKLLAKRC	MPHSTVHSHTLISDDSVLLFPDLISYCAYPLRVNPYGRSVADDSERWLLNGAHLSDKKRKAFLRLRAGDLASMCYPDASAKSLRVVADYMNYLFKLDDWTDEFEAEDVDGMRDCVLAALRDPLHYETDKAVGKLAKSFFGRFVQHGGPLRTKRFIDTMVLFFRAVRQQALDRTYDDIPDLESYIALRRDTSGCKPCFALIEFAGGYDLPDAVVEHPSIQILQDATNDLVTWSNDIFSYNVEQSRGDTHNMVVVLMHEQGLAVQEAIDAVADLCERSIDTFEQTRRSLPSWGPIVDSNVESYIDGLQNWIIGSLHWSFLTERYFGKDRRDVKRKLFVKLLAKRC
ShSTS8	MPAIIRQFILPDLFALSSAFPDATNPHWKRACTESRDWVNSYRVFSDEERRAFFTQGQSELLCSHAYPYAGYEQFRTCCDFINLLFVL**DEISD**EQTADGAWATGRIFLQVLQDPEWDDGSKVAQMTRDLRARVVSTGVKPHTFRRLVHMCRDYIASVVEEAGLRERGEVLDIESYIELRRNNSAVLTCFALIPYILGIDLPDEVVNDPNFSALNLAAVDMVCWA**NDIYSYDME**QAKGLEGNNIMTVLTEMHGLTMQDASDYVGEQYKALMDLFLDNQAALRSFGPSVDADVRRYVDAVRHWPRGNLSWSFETPRYFREKREEIENTRVVILRPRPEPKMKSQN	MPAIIRQFILPDLFALSSAFPDATNPHWKRACTESRDWVNSYRVFSDEERRAFFTQGQSELLCSHAYPYAGYEQFRTCCDFINLLFVLDEISDEQTADGAWATGRIFLQVLQDPEWDDGSKVAQMTRDLRARVVSTGVKPHTFRRLVHMCRDYIASVVEEAGLRERGEVLDIESYIELRRNNSAVLTCFALIPYILGIDLPDEVVNDPNFSALNLAAVDMVCWANDIYSYDMEQAKGLEGNNIMTVLTEMHGLTMQDASDYVGEQYKALMDLFLDNQAALRSFGPSVDADVRRYVDAVRHWPRGNLSWSFETPRYFREKREEIENTRVVILRPRPEPKMKSQN	MPAIIRQFILPDLFALSSAFPDATNPHWKRACTESRDWVNSYRVFSDEERRAFFTQGQSELLCSHAYPYAGYEQFRTCCDFINLLFVLDEISDEQTADGAWATGRIFLQVLQDPEWDDGSKVAQMTRDLRARVVSTGVKPHTFRRLVHMCRDYIASVVEEAGLRERGEVLDIESYIELRRNNSAVLTCFALIPYILGIDLPDEVVNDPNFSALNLAAVDMVCWANDIYSYDMEQAKGLEGNNIMTVLTEMHGLTMQDASDYVGEQYKALMDLFLDNQAALRSFGPSVDADVRRYVDAVRHWPRGNLSWSFETPRYFREKREEIENTRVVILRPRPEPKMKSQN
ShSTS10	MGATSTTQIPHSTSLESYILPDLLRLSSPFKASMNPHWATTASESSAWFSSYSIFSDQELTEFAVCKVELLVAYAYPHANYETFRTCCDFMNLIFAI**DEISD**MQNGEDAGETGDVFLNALRDAEWTDGSALAEMTRDFRARFLRSAGPQSFRRFLKVSEDYIDCVTKEAGYRERSQILDMESFKHLRRDNSGVPLMLGLLEYTLGIDLPDVVFEDTALSRIYWAAVDMVWWA**NDVYSYKVE**QAKGLAGNNILSVLMAAKNIDLQDASDYVGDVYAGLMKEYTEAKAELVSKSFGSKELDAAVKKYVDKMENWPIGNLEWSFASMRYFGSQSGEIKRTRTVMLRPQDVLA	MGATSTTQIPHSTSLESYILPDLLRLSSPFKASMNPHWATTASESSAWFSSYSIFSDQELTEFAVCKVELLVAYAYPHANYETFRTCCDFMNLIFAIDEISDMQNGEDAGETGDVFLNALRDAEWTDGSALAEMTRDFRARFLRSAGPQSFRRFLKVSEDYIDCVTKEAGYRERSQILDMESFKHLRRDNSGVPLMLGLLEYTLGIDLPDVVFEDTALSRIYWAAVDMVWWANDVYSYKVEQAKGLAGNNILSVLMAAKNIDLQDASDYVGDVYAGLMKEYTEAKAELVSKSFGSKELDAAVKKYVDKMENWPIGNLEWSFASMRYFGSQSGEIKRTRTVMLRPQDVLA	MGATSTTQIPHSTSLESYILPDLLRLSSPFKASMNPHWATTASESSAWFSSYSIFSDQELTEFAVCKVELLVAYAYPHANYETFRTCCDFMNLIFAIDEISDMQNGEDAGETGDVFLNALRDAEWTDGSALAEMTRDFRARFLRSAGPQSFRRFLKVSEDYIDCVTKEAGYRERSQILDMESFKHLRRDNSGVPLMLGLLEYTLGIDLPDVVFEDTALSRIYWAAVDMVWWANDVYSYKVEQAKGLAGNNILSVLMAAKNIDLQDASDYVGDVYAGLMKEYTEAKAELVSKSFGSKELDAAVKKYVDKMENWPIGNLEWSFASMRYFGSQSGEIKRTRTVMLRPQDVLA
ShSTS11	MATSNPPSTPHHESLILPDLLSLSTPFNGSTNPHWAMAAPESSAWVSSYNLFSDRKRTDFITGSNELLVSHTYPHADYDAFRTCCDFVNLLFVI**DEISD**DQSGKAARRTGEVYLNAMRDPEWTDGSDLAKMTQQFRARFLRSVGPQSFRRFLRHSEDYIDCVAKEAEYRERGQVLDMDSFKSLRRENSAIRLCFGLFEFTLGIDLPDSVFEDETFMKMYWASADMVCWA**NDVYSYNVE**QAKGHSGNNIVTVLMAARDIDMQAASDYVGEYYAELMEEYMTAKAELASKSFGSRDLDEDVWKYVNAMENWPIGNLEWSFKTNRYFGTLHDEVKRTRLVVIKPRKVVV	MATSNPPSTPHHESLILPDLLSLSTPFNGSTNPHWAMAAPESSAWVSSYNLFSDRKRTDFITGSNELLVSHTYPHADYDAFRTCCDFVNLLFVIDEISDDQSGKAARRTGEVYLNAMRDPEWTDGSDLAKMTQQFRARFLRSVGPQSFRRFLRHSEDYIDCVAKEAEYRERGQVLDMDSFKSLRRENSAIRLCFGLFEFTLGIDLPDSVFEDETFMKMYWASADMVCWAN**VSDTTHTPHSLPSSPINLRHNGS**KGHSGNNIVTVLMAARDIDMQAASDYVGEYYAELMEEYMTAKAELASKSFGSRDLDEDVWKYVNAMENWPIGNLEWSFKTNRYFGTLHDEVKRTRLVVIKPRKVVV	*(MATSNPPSTPHHESLILPDLLSLSTPFNGSTNPHWA)*MAAPESSAWVSSYNLFSDRKRTDFITGSNELLVSHTYPHADYDAFRTCCDFVNLLFVIDEISDDQSGKAARRTGEVYLNAMRDPEWTDGSDLAKMTQQFRARFLRSVGPQSFRRFLRHSEDYIDCVAKEAEYRERGQVLDMDSFKSLRRENSAIRLCFGLFEFTLGIDLPDSVFEDETFMKMYWASADMVCWANDVYSYNVEQAKGHSGNNIVTVLMAARDIDMQAASDYVGEYYAELMEEYMTAKAELASKSFGSRDLDEDVWKYVNAMENWPIGNLEWSFKTNRYFGTLHDEVKRTRLVVIKPRKVVV

Table S5. Continued.

Gene name	Active protein sequence	Predicted sequence by Augustus (*A. oryzae*)	Predicted sequence by Augustus (*C. cinereus*)
ShSTS12	MQTPPSCESYILPDLLHISQPFKASTNPHWLKAAPESSAWVSSYRIFTDRKRMEFILGSNELLASHTYPQANYDTFRTCCDFINLLFVI**DEISD**DQSGEGARLTGKVFLNAMRDPECTDGSILAKMTRDFRARLLQSIGAQSFRRFLKILEDYIESVTKEAEYREHGYVLDMEQFKILRRQNSAVRCAFGMVEHTLGIDLPDVVFEDATFMKVYWAAIDMVWCA**NDVYSYSME**LAKGDNHSANNIMSVFMAANGVSLQAASDYVGRHYARLMEEYLSARAELASKSFGSQELDSDVLKYIDAMENWPIGNIEWSFKTSRYFGPQNDEVKQTLVVKLKPQEVPA	MQTPPSCESYILPDLLHISQPFKASTNPHWLKAAPESSAWVSSYRIFTDRKRMEFILGSNELLASHTYPQANYDTFRTCCDFINLLFVIDEISDDQSGEGARLTGKVFLNAMRDPECTDGSILAKMTRDFRARLLQSIGAQSFRRFLKILEDYIESVTKEAEYREHGYVLDMEQFKILRRQNSAVRCAFGMVEHTLGIDLPDVVFEDATFMKVYWAAIDMVWCAN**VNMFPPPGYLIQ**DVYSYSMELAKGDNHSANNIMSVFMAANGVSLQAASDYVGRHYARLMEEYLSARAELASKSFGSQELDSDVLKYIDAMENWPIGNIEWSFKTSRYFGPQNDEVKQTLVVKLKPQEVPA	MQTPPSCESYILPDLLHISQPFKASTNPHWLKAAPESSAWVSSYRIFTDRKRMEFILGSNELLASHTYPQANYDTFRTCCDFINLLFVIDEISDDQSGEGARLTGKVFLNAMRDPECTDGSILAKMTRDFRARLLQSIGAQSFRRFLKILEDYIESVTKEAEYREHGYVLDMEQFKILRRQNSAVRCAFGMVEHTLGIDLPDVVFEDATFMKVYWAAIDMVWCAN**VNMFPPPGYLIQ**DVYSYSMELAKGDNHSANNIMSVFMAANGVSLQAASDYVGRHYARLMEEYLSARAELASKSFGSQELDSDVLKYIDAMENWPIGNIEWSFKTSRYFGPQNDEVKQTLVVKLKPQEVPA
ShSTS13	MAATSTPLSCKPEDDGYSVPMTVPYRMPDLFTAWPWKTTRNPLYLEAKADSDGWLATFGDFFQKYKDCILDCDGTLFCAYGLPSTDKERLRCAGDVNTLLFVV**DDILD**YEDDATVKERMRILMDAMENPFKPRPDGEWIVGEMTRQFWERTIKVSNTLSQERFMEGVRDCMAGMVRESLARKASRQHDSIHGFLKTRRDSAGCNLVFALSELEVAVPAEAMKHSQIRELIAIIQDISCIA**NDTFSYRRE**QSRGRTEDNIITVVMNNLGTDIPGAIDWVENHHKELMKEFIEKHDKVPKWGEPTDSKVKNYIESLAGWVWANARWSFECRRYMGDEGMEIMEKSRWVRAIPKERI	MAATSTPLSCKPEDDGYSVPMTVPYRMPDLFTAWPWKTTRNPLYLEAKADSDGWLATFGDFFQKYKDCILDCDGTLFCAYGLPSTDKERLRCAGDVNTLLFVVDDILDYEDDATVKERMRILMDAMENPFKPRPDGEWIVGEMTRQFWERTIKVSNTLSQERFMEGVRDCMAGMVRESLARKASRQHDSIHGFLKTRRDSAGCNLVFALSELEVAVPAEAMKHSQIRELIAIIQDISCIANDTFSYRREQSRGRTEDNIITVVMNNLGTDIPGAIDWVENHHKELMKEFIEKHDKVPKWGEPTDSKVKNYIESLAGWVWANARWSFECRRYMGDEGMEIMEKSRWVRAIPKERI	MAATSTPLSCKPEDDGYSVPMTVPYRMPDLFTAWPWKTTRNPLYLEAKADSDGWLATFGDFFQKYKDCILDCDGTLFCAYGLPSTDKERLRCAGDVNTLLFVVDDILDYEDDATVKERMRILMDAMENPFKPRPDGEWIVGEMTRQFWERTIKVSNTLSQERFMEGVRDCMAGMVRESLARKASRQHDSIHGFLKTRRDSAGCNLVFALSELEVAVPAEAMKHSQIRELIAIIQDISCIANDTFSYRREQSRGRTEDNIITVVMNNLGTDIPGAIDWVENHHKELMKEFIEKHDKVPKWGEPTDSKVKNYIESLAGWVWANARWSFECRRYMGDEGMEIMEKSRWVRAIPKERI
ShSTS16	MAVATSVATPVPTPAYSAGRAPAKEKKIYLPDTLAEWPWPRAINPHYAEAKEESQAWAASFNAFSPKAQHAFNRCDFNLLASLAYPLATKHGCRSGCDLMNLFFVI**DEYSD**IAPVEEVRQQKDIVMDALRNPHKPRPEGEWVGGEVARQFWALTITNASAQSQKHFIETFDEYLDSVVQQAEDRSESRIRDIQSYIDVRRNTIGAKPSFALLELDMDLPDEVLAHPTIQSLSLATIDMLCLG**NDIVSYNLE**QARGDASHNIITIVMNELNLDVNGAMRWVGDFHKQLEKQFFEAFNNLPKWGNAELDAQIAVYCDGLGNWVRANDQWSFESERYFGARGLEIMETKTLAMMPIQRTEALGPQLVDDSIL	MAVATSVATPVPTPAYSAGRAPAKEKKIYLPDTLAEWPWPRAINPHYAEAKEESQAWAASFNAFSPKAQHAFNRCDF**SKSFINFRSLFFAYLLTVYVFVRAFAQTFLHPSLTRSPLNYLPRFFSPD**GCRSGCDLMNLFFVIDEYSDIAPVEEVRQQKDIVMDALRNPHKPRPEGEWVGGEVARQFWALTITNASAQSQKHFIETFDEYLDSVVQQAEDRSESRIRDIQSYIDVRRNTIGAKPSFALLELDMDLPDEVLAHPTIQSLSLATIDMLCLGNDIVSYNLEQARGDASHNIITIVMNELNLDVNGAMRWVGDFHKQLEKQFFEAFNNLPKWGNAELDAQIAVYCDGLGNWVRANDQWSFESERYFGARGLEIMETKTLAMMPIQRTEALGPQLVDDSIL	MAVATSVATPVPTPAYSAGRAPAKEKKIYLPDTLAEWPWPRAINPHYAEAKEESQAWAASFNAFSPKAQHAFNRCDFN*(LLASLAYPLATKH)*GCRSGCDLMNLFFVIDEYSDIAPVEEVRQQKDIVMDALRNPHKPRPEGEWVGGEVARQFWALTITNASAQSQKHFIETFDEYLDSVVQQAEDRSESRIRDIQSYIDVRRNTIGAKPSFALLELDMDLPDEVLAHPTIQSLSLATIDMLCLGNDIVSYNLEQARGDASHNIITIVMNELNLDVNGAMRWVGDFHKQLEKQFFEAFNNLPKWGNAELDAQIAVYCDGLGNWVRANDQWSFESERYFGARGLEIMETKTLAMMPIQRTEALGPQLVDDSIL
ShSTS17	MSQTYTIPDTLANWPWKRKINQHYEEVKMESASWARSFHAFSPQAQDAFDRCDFNLLASLAYPLADKQRLRTGCDLMNLFFVI**DEHSD**LSSVADAETQAQIIMNALMNPEKPRPHGEWVGGEVARQYWELAIKTATPKSQRRFVAAFDDYMNAVVQQAKDRTHSTIRDIDSYMEVRRKTIGAWPSFALLELDMDLPEDFMDHPVMHELHVLSISMICLG**NDIVSWNLE**QSRGDDTHNIVRIVMNQLDTDINGAMAWVEMYHKELEVKFMDIFTKSQEWNKSMNKDISRYVEGLGNWVRANDQWSFESKRYFGDRGLEIMSKRTVSMMPKRNDVLNGTQLDIGPVIVDGSIL	*(MSQTYTIPDTLANWPWKRKINQHYEEVKMESASWARSFHAFSPQAQDAFDRCDFNLLASLAYPLADKQRLRTGCDL)*MNLFFVIDEHSDLSSVADAETQAQIIMNALMNPEKPRPHGEWVGGEVARQYWELAIKTATPKSQRRFVAAFDDYMNAVVQQAKDRTHSTIRDIDSYMEVRRKTIGAWPSFALLELDMDLPEDFMDHPVMHELHVLSISMICLGNDIVSWNLEQSRGDDTHNIVRIVMNQLDTDINGAMAWVEMYHKELEVKFMDIFTKSQEWNKSMNKDISRYVEGLGNWVRANDQWSFESKRYFGDRGLEIMSKRTVSMMPKRNDVLNGTQLDIGPVIVDGSIL	MSQTYTIPDTLANWPWKRKINQHYEEVKMESASWARSFHAFSPQAQDAFDRCDFNLLASLAYPLADKQRLRTGCDLMNLFFVIDEHSDLSSVADAETQAQIIMNALMNPEKPRPHGEWVGGEVARQYWELAIKTATPKSQRRFVAAFDDYMNAVVQQAKDRTHSTIRDIDSYMEVRRKTIGAWPSFALLELDMDLPEDFMDHPVMHELHVLSISMICLGNDIVSWNLEQSRGDDTHNIVRIVMNQLDTDINGAMAWVEMYHKELEVKFMDIFTKSQEWNKSMNKDISRGASRASVILEIEA
ShSTS18	MTVVDSPQRFCIPNCLEYWPWPRHINPHYQEVKKASAAWAESFGAFNPKAQHAYNACDFNLLASLAYPLESEERLRTGCDLMNMFFVF**DEYSD**VSSPKDVIQQAAIIMDALRNPYAPRLEDEWVGGEVTRQFWKRAIKTATAGAQRRFIDAFESYTQSVVQQAKDRHHGFIRDVDSYLEMRRETIGAKPSFVVLEMDMTLPDEVLAHPVIQQLSALSTDMICLG**NDICSYNVE**QARGDDLHNIITIAMNQFDIDIAGAMDWVVKYHAKLERKFLHLYNNCLPSWGKELDPQVERYVCGLGNWVRASDQWGFESERYFGKKGKEIFKRRWVNLMQPERAQDVGPTLVDGTRL	MTVVDSPQRFYIPNCLEYWPWPRHINPHYQEVKKASAAWAESFGAFNPKAQHAYNACDFNLLASLAYPLESEERLRTGCDLMNMFFVFDEYSDVSSPKDVIQQAAIIMDALRNPYAPRPDDEWVGGEVTRQFWKRAIKTATAGAQRRFIDAFESYTQSVVQQAKDRHHGFIRDVDSYLEMRRETIGAKPSFVVLQMDMTLPDEVLAHPVIQQLSALSTDMICLGNDICSYNVEQARGDDLHNIITIAMNQFDIDIAGAMDWVVKYHAKLERKFLYLYNNGLPSWGKELDPQVERYVCGLGNWVRASDQWGFESERYFGKKGKEIFKRRWVNLMQPERAQDIGPTLVDGTRL	MTVVDSPQRFYIPNCLEYWPWPRHINPHYQEVKKASAAWAESFGAFNPKAQHAYNACDFNLLASLAYPLESEERLRTGCDLMNMFFVFDEYSDVSSPKDVIQQAAIIMDALRNPYAPRPDDEWVGGEVTRQFWKRAIKTATAGAQRRFIDAFESYTQSVVQQAKDRHHGFIRDVDSYLEMRRETIGAKPSFVVLQMDMTLPDEVLAHPVIQQLSALSTDMICLGNDICSYNVEQARGDDLHNIITIAMNQFDIDIAGAMDWVVKYHAKLERKFLYLYNNGLPSWGKELDPQVERYVCGLGNWVRASDQWGFESERYFGKKGKEIFKRRWVNLMQPERAQDIGPTLVDGTRL

Table S6. Oligonucleotides used for construction of expression plasmids. Sequences with bold characters are homologous sequences to those of the vector. Sequences with italic characters with underline are homologous sequences to those of the linker.

Insert	Sequence 5'-3'	Size Vector
CpSTS1	F: **GCAAGCTCCGGAATTC**ATGCCTGCTGCTATCCCTAAG	1297 bp
	R: **TACCGAGCTCGAATTC**TCAAGCGGCGTTTTCGTCGAC	pTAex3-EcoRI
CpSTS2	F: **CCGAATTCGAGCTCGGTACC**ATGATGCAATTCTACCTTCC	1252 bp
	R: **ACTACAGATCCCCGGGTACC**CTATTCGAAGTCTGAATCGG	pUARA2-KpnI
CpSTS3	F: **CCGAATTCGAGCTCGGTACC**ATGGTCGCCACTACAGCA	1341 bp
	R: **ACTACAGATCCCCGGGTACC**TTAGGCCTTGAGTGGCAAG	pUARA2-KpnI
CpSTS4	F: **GCAAGCTCCGGAATTC**ATGTCCGAGTCCAACCAACAC	1414 bp
	R: **TACCGAGCTCGAATTC**TCAGTCTAGTTGGTGTAGAGG	pTAex3-EcoRI
CpSTS5	F: **CCGAATTCGAGCTCGGTACC**ATGTCTACCATAGCCACAG	1212 bp
	R: **ACTACAGATCCCCGGGTACC**TTACGCACGCTGTGGGAG	pDP1031-KpnI
CpSTS6	F: **CCGAATTCGAGCTCGGTACC**ATGCTCACTACGCCTTACT	1349 bp
	R: **ACTACAGATCCCCGGGTACC**TCATACTATAATCCCAGCCG	pDP1031-KpnI
CpSTS7	F: **CCGAATTCGAGCTCGGTACC**ATGACTACCAGCGGATTCC	1721 bp
	R: **ACTACAGATCCCCGGGTACC**CTAATCCTCACCGGGATC	pDP1031-KpnI
CpSTS8	F: **CCGAATTCGAGCTCGGTACC**ATGTCCAACACCTTCATTGT	1368 bp
	R: **ACTACAGATCCCCGGGTACC**TTACATGATGGACGCGAAC	pDP1031-KpnI
CpSTS9	F: **CCGAATTCGAGCTCGGTACC**ATGACGGTACTTCAATTTGTG	1282 bp
	R: **ACTACAGATCCCCGGGTACC**TTATTCCACAAGTTTCACAGT	pDP1031-KpnI
CpSTS10	F: **CCGAATTCGAGCTCGGTACC**ATGGCTCCAAGCTATCAGC	1277 bp
	R: **ACTACAGATCCCCGGGTACC**TCAAGGCAACACATTGTCG	pDP1031-KpnI
CpSTS11	F: **CCGAATTCGAGCTCGGTACC**ATGTCCCAGAAATCTCCTTC	1262 bp
	R: **ACTACAGATCCCCGGGTACC**CTAATCGGCGAGATCGAC	pDP1031-KpnI
CpSTS12	F: **CCGAATTCGAGCTCGGTACC**ATGGAAAGCCTGCCAGTAA	1267 bp
	R: **ACTACAGATCCCCGGGTACC**TCATAAGTAATTCTTCCGACC	pDP1031-KpnI
CpSTS13	F: **CCGAATTCGAGCTCGGTACC**ATGCCCCAAGACCCGCCT	1285 bp
	R: **ACTACAGATCCCCGGGTACC**CTAGTCTGAAATATGGACGT	pUSA2-KpnI
CpSTS14	F: **CCGAATTCGAGCTCGGTACC**ATGAACGTAATGAACACTCC	1235 bp
	R: **ACTACAGATCCCCGGGTACC**TCAGCTCAAGCTCGTAGTT	pDP1031-KpnI
CpSTS15	F: **CCGAATTCGAGCTCGGTACC**ATGTCATCCCACCGACCC	1103 bp
	R: **ACTACAGATCCCCGGGTACC**CTACAGCCGATTATATCTTC	pDP1031-KpnI
CpSTS16	F: **CCGAATTCGAGCTCGGTACC**ATGCCTCTCCCAGTAACG	1285 bp
	R: **ACTACAGATCCCCGGGTACC**TTACTCATTCAAGTTGACACT	pUSA2-KpnI
CpSTS17	F: **CCGAATTCGAGCTCGGTACC**ATGAACTCAGACCTGCAATC	1102 bp
	R: **ACTACAGATCCCCGGGTACC**TTAGGCAGCTACGGCGCG	pDP1031-KpnI
CpSTS18	F: **CCGAATTCGAGCTCGGTACC**ATGGTCCACTGGAATCACC	1625 bp
	R: **ACTACAGATCCCCGGGTACC**TCATGAATCACCGTACAGC	pDP1031-KpnI
ShSTS1	F: **CAAGCTCCGAATTCGAGCTCGGTACC**ATGGAGTCCGTCAGGGAAC	1236 bp
	R: **ATCCCCGGGTACC***TTACTCAAACGTG*CTAAGCTTTATG	pDP1032-KpnI/SpeI
linker	F: *CACGTTTGAGTAA*GGTACCCGGGGATCTGTAG	869 bp
	R: *CCACGACTGTCAT*CTAGTGCGGCCGCTAGCT	pDP1032-KpnI/SpeI

Insert	Sequence 5’-3’	Size Vector
ShSTS18	F: **GCGGCCGCACTAG***ATGACAGTCGTGG*ACAGC	1292 bp
	R: **TCATCGGGTACGACTACCCGGGTCA**TCATAATCTGGTCCCATCAAC	pDP1032-KpnI/SpeI
ShSTS3	F: **CCGAATTCGAGCTCGGTACC**ATGTCTACTGCGAAACCTG	1333 bp
	R: **ACTACAGATCCCCGGGTACC**CTAGGGCACAGTCACCTC	pDP1031-KpnI
ShSTS4	F: **CCGAATTCGAGCTCGGTACC**ATGTCTGAAACCAAAGTTGG	3335 bp
	R: **ACTACAGATCCCCGGGTACC**TCAATGGGTGACGGTATAC	pDP1031-KpnI
ShSTS5	F: **CCGAATTCGAGCTCGGTACC**ATGGAACTCTCTTCACTGC	1493 bp
	R: **ACTACAGATCCCCGGGTACC**TCAGGAGCAATCCTTGCTG	pDP1031-KpnI
ShSTS7	F: **CCGAATTCGAGCTCGGTACC**ATGCCTCATTCCACCGTG	1273 bp
	R: **ACTACAGATCCCCGGGTACC**TCAACACCTCTTCGCGAG	pDP1031-KpnI
ShSTS8	F: **CCGAATTCGAGCTCGGTACC**ATGCCTGCTATCATCAGAC	1293 bp
	R: **ACTACAGATCCCCGGGTACC**TCAATTCTGGCTCTTCATCT	pDP1031-KpnI
ShSTS10	F: **CCGAATTCGAGCTCGGTACC**ATGGGAGCGACTTCGACG	1250 bp
	R: **ACTACAGATCCCCGGGTACC**CTACGCAAGAACATCTTGAG	pDP1031-KpnI
ShSTS11	F: **CAAGCTCCGAATTCGAGCTCGGTACC**ATGGCTACTTCAAATCCTC	1296 bp
	R: **ATCCCCGGGTACC***CTAAACAACGACT*TTCCG	pDP1032-KpnI/SpeI
linker	F: *AGTCGTTGTTTAG*GGTACCCGGGGATCTGTAG	869 bp
	R: *TAGCGACAGCCAT*CTAGTGCGGCCGCTAGCT	pDP1032-KpnI/SpeI
ShSTS16	F: **GCGGCCGCACTAG***ATGGCTGTCGCTA*CCTCTG	1447 bp
	R:**TCATCGGGTACGACTACCCGGGTCA**TCACAAAATGGAGTCATCAACAAG	pDP1032-KpnI/SpeI
ShSTS12	F: **CCGAATTCGAGCTCGGTACC**ATGCAGACTCCACCCTCT	1256 bp
	R: **ACTACAGATCCCCGGGTACC**CTAGGCAGGAACTTCCTG	pDP1031-KpnI
ShSTS13	F: **CCGAATTCGAGCTCGGTACC**ATGGCAGCTACGTCCACC	1409 bp
	R: **ACTACAGATCCCCGGGTACC**TTATATACCATCTAATCCGTG	pDP1031
ShSTS17	F: **CAAGCTCCGAATTCGAGCTCGGTACC**ATGTCGCAGACATATACAATAC	1322 bp
	R: **ACTACAGATCCCCGGGTACC**TTACAGAATGGACCCATCG	pDP1031-KpnI
CpSTS5-Fr1	F: **GTGCATCATCATCATCATCATATCGAAGGTAGGCAT**ATGTCTACCATAGCCAC	607 bp
	R: *TTGTAGGCGT*ATTCGATAAGAGCCCAGC	pColdI-NdeI
CpSTS5-Fr2	F: *CTTATCGAAT*ACGCCTACAACCTGCACATTC	428 bp
	R: **CTCGAGGGTACCGAGCTCCATATG**TTACGCACGCTGTGGGAG	pColdI-NdeI
CpSTS7-Fr1	F: **CGGTACCCTCGAGGGATCCGAATTC**ATGACTACCAGCGGATTC	664 bp
	R: *GAGAATACATCATTGGCAAT*GATCAGCATATC	pColdI-EcoRI
CpSTS7-Fr2	F: *ATTGCCAATGATGTATTCTC*ATTCAAAGTCGAG	464 bp
	R: **TTACCTATCTAGACTGCAGGTCGAC**CTATCGCCGAAGGAGAGTAG	pColdI
CpSTS8-Fr1	F: **AAGGTAGGCATATGGAGCTCGGTACC**ATGTCCAACACCTTCATTG	140 bp
	R: *ATAGCAAATTGCC*AATATCCATAGACGTCCTG	pColdI-kpnI/BamHI
CpSTS8-Fr2	F: *GTCTATGGATATT*GGCAATTTGCTATGGCTTTATG	540bp
	R: *AAGATCTAGAAAC*TCTTTAACTAGCCGGCAG	pColdI-kpnI/BamHI
CpSTS8-Fr3	F: *GGCTAGTTAAAGA*GTTTCTAGATCTTCATGTCCG	112 bp
	R: *CTCAAACGACCAT*CTGTCGTTGCCAATGACAG	pColdI-kpnI/BamHI
CpSTS13-Fr1	F: **AAGGTAGGCATATGGAGCTCGGTACC**ATGCCCCAAGACCCGCCT	134 bo
	R: *GTAGACATGTTTT*GAATACCGTAGAATCCACTCTCTTGCC	pColdI-kpnI/BamHI

Insert	Sequence 5’-3’	Size Vector
CpSTS13-Fr2	F: *TTCTACGGTATTC*AAAACATGTCTACGGACC	258 bp
	R: *CCTTCTACGAAAA*TCTTGAAGTATTGCGCAAAG	pColdI-kpnI/BamHI
CpSTS13-Fr3	F: *CAATACTTCAAGA*TTTTCGTAGAAGGAACCTC	652 bp
	R: **CTGCAGGTCGACAAGCTTGAATTCGGATCC**CTAGTCTGAAATATGGACG	pColdI-kpnI/BamHI
CpSTS14-Fr1	F: **AGGCATATGGAGCTCGGTACC**ATGAACGTAATGAACACTCCCATC	623 bp
	R: *ATCGAGAGATAGG*AAGACACCGGCTTCCACC	pColdI-KpnI/HindIII
CpSTS14-Fr2	F: *CCGGTGTCTT*CCTATCTCTCGATCTCGC	484 bp
	R: **ATCTAGACTGCAGGTCGACAAGCTT**TCAGCTCAAGCTCGTAGTTC	pColdI-KpnI/HindIII
CpSTS16-Fr1	F: **GTGCATCATCATCATCATCATATCGAAGGTAGGCAT**ATGCCTCTCCCAGTAAC	246 bp
	R: *ATACATTATG*CAGATCCATGACAAAACG	pColdI-NdeI
CPSTS16-Fr2	F: *CATGGATCTG*CATAATGTATCATGGCTTTTC	798 bp
	R: **CTCGAGGGTACCGAGCTCCATATG**TTACTCATTCAAGTTGACAC	pColdI-NdeI
CpSTS17-Fr1	F: **GAGCTCGGTACCCTCGAGGGATCCGAATTC**ATGAACTCAGACCTGCAATC	297 bp
	R: *CAAGTTTGGTAGCCTTGTTCGCCTCC*AGACC	pColdI-EcoRI
CpSTS17-Fr2	F: *GGAGGCGAACAAGGCTACCAAACTTG*TTGATCG	263 bp
	R: *ACTGGTAGAGAACCAGAGGCCGACGT*CTAATTC	pColdI-EcoRI
CpSTS17-Fr3	F: *ACGTCGGCCTCTGGTTCTCTACCAGT*CTTTTTCTATGGGG	433 bp
	R: **AGAGATTACCTATCTAGACTGCAGGTCGAC**TTAGGCAGCTACGGCGCG	pColdI-EcoRI
ShSTS3-Fr1	F: **AAGGTAGGCATATGGAGCTCGGTACC**ATGTCTACTGCGAAACCTG	461 bp
	R: CATCCTCTGCACTCGAAGATACTCCATCATGTC	pColdI
ShSTS3-Fr2	F: TGGAGTATCTTCGAGTGCAGAGGATGGCACC	448 bp
	R: CATATCTCGTAGAATGATGATGCCACCAAGTATTCC	pColdI
ShSTS3-Fr3	F: GTGGCATCATCATTCTACGAGATATGCAGAG	242 bp
	R: **TTGAATTCGGATCCCTCGAGGGTACC**ATGAACCCGGATCCTTTC	pColdI
ShSTS5-Fr1	F: **AGGCATATGGAGCTCGGTACC**ATGGAACTCTCTTCACTGCG	416 bp
	R: *ATTAAACAGGCTCTGGAGCA*TAGCCGCG	pColdI-KpnI
ShSTS5-Fr2	F: *TGCTCCAGAGCCTGTTTAAT*CGTATCAAGACG	689 bp
	R: **TTACCTATCTAGACTGCAGGTCGAC**TCAGGAGCAATCCTTGCTG	pColdI-KpnI
ShSTS10-Fr1	F: **GAGCTCGGTACCCTCGAGGGATCCGAATTC**ATGGGAGCGACTTCGACG	410 bp
	R: *TCGGGCACGAAAG*TCCCTTGTCATTTCCGCG	pColdI-EcoRI/SalI
ShSTS10-Fr2	F: *AAATGACAAGGGA*CTTTCGTGCCCGATTCCTG	289 bp
	R: *ATGAGTACACATC*GTTCGCCCACCATACCATATC	pColdI-EcoRI/SalI
ShSTS10-Fr3	F: *ATGGTGGGCGAAC*GATGTGTACTCATATAAGGTG	351 bp
	R: **AGAGATTACCTATCTAGACTGCAGGTCGAC**CTACGCAAGAACATCTTG	pColdI-EcoRI/SalI
ShSTS12-Fr1	F: **AGGCATATGGAGCTCGGTACC**ATGCAGACTCCACCCTCTTG	386 bp
	R: *ACGCGCGCGAAAGTCCCTTGTCA*TCTTCGCTAAAATAG	pColdI-KpnI
ShSTS12-Fr2	F: *TGACAAGGGACTTTCGCGCGCGT*CTTCTTC	646 bp
	R: **GGTCGACAAGCTTGAATTCGGATCC**CTAGGCAGGAACTTCCTGAGG	pColdI-KpnI
ShSTS13-Fr1	F: **AGGCATATGGAGCTCGGTACC**ATGGCAGCTACGTCCACC	262 bp
	R: *CGTAGTCGTT*CTTTGTCAGTCGATGGAAGC	pColdI-KpnI/SalI
ShSTS13-Fr2	F: *ACTGACAAAG*AACGACTACGATGCGCAG	847 bp
	R: **TTACCTATCTAGACTGCAGGTCGAC**TTATATACCATCTAATCCGTGAATC	pColdI-KpnI/SalI

Insert	Sequence 5’-3’	Size Vector
ShSTS18-Fr1	F: **AGGCATATGGAGCTCGGTACC**ATGACAGTCGTGGACAGC	178 bp
	R: *GCCAACAAGT*TGAAGTCGCAGGCATTATAC	pColdI-KpnI/BamHI
ShSTS18-Fr2	F: *TGCGACTTCA*ACTTGTTGGCTTCGCTGG	878 bp
	R: **GGTCGACAAGCTTGAATTCGGATCC**TCATAATCTGGTCCCATCAAC	pColdI-KpnI/BamHI
SC103-up	F: TTGTAAAACGACGGCCAGTGATTAATTTTTTTCGTTTAAGTCTGTC	
	R: *GAATGAACCCGGATCCGGAGAATCAG*TCCAAGAAC	pUC19-HindIII/EcoRI
SC103-down	F: *CTGATTCTCCGGATCCGGGTTCATTC*TTTCGGTTC	
	R: CTATGACCATGATTACGCCAGTAAGAGCTTGTTGGGAG	pUC19-HindIII/EcoRI
PamyB-MCS-TamyB	F: **CTGATTCTCCGGATCC**CCATCATGGTGTTTTGATCA	825 bp
	R: **GAATGAACCCGGATCC**TTTCCTATAATAGACTAGCGT	pUC19-SC103-BamHI
PamyB-MCS-TamyB-PamyB-MCS-TamyA	F: **CTGATTCTCCGGATC**TCATGGTGTTTTGATCATTTTA	1752 bp
	R: **GAATGAACCCGGATC**GGTGGAAGAGTCTTAACGT	pUC19-SC103-BamHI
U6P-SC103	F: TCGAGCTCGGTACCCTAATGCCGGCTCATTCAAAC	
	R: TATAGAAGGAAGGGTAAAGCACTTGTTCTTCTTTACAATG	ppAsACas9-SmaI
SC103-U6T	F: GCTTTACCCTTCCTTCTATAGTTTTAGAGCTAGAAATAGC	
	R: CAATTGCCCGATCCCCAGCAGCTCTATATCACGTGACG	ppAsACas9-SmaI
U6&-SC103-U6T	F: TCGAGCTCGGTACCCTAATGCCGGCTCATTCAAAC	
	R: CAATTGCCCGATCCCCAGCAGCTCTATATCACGTGACG	ppAsACas9-SmaI
ple3	F: **CAGCAAGCTCCGAATTCGAGCTCGGTAC**ATGGGTCTATCCGAAGATC	3039 bp
	R: *ATCCCCGGGTAC*TCAGTGGTGGATTCCATTGC	pDP1032-KpnI/SpeI
ple4	F: *CGGCCGCACTAGT*ATGAGAATACCTAACATCTTTC	1285 bp
	R: **TCATCGGGTACGACTACCCGGGTCACTAGT**CTACTCTGCAATGTACAAC	pDP1032-KpnI/SpeI
linker (ple3-ple4)	F: AATCCACCACTGA*GTACCCGGGGAT*CTGTAGTAG	
	R: TAGGTATTCTCAT*ACTAGTGCGGCCG*CTAGCTC	pDP1032-KpnI/SpeI
ple5	F: **CCGAATTCGAGCTCG**ATGCTATCTGTCGACCTTC	2297 bp
	R: **ACTACAGATCCCCGG**CTACAATGCAGCGAACGC	pDP1031-KpnI
ple6	F: **ATCGATTTGAGCTAG**ATGAATCTTTCTGAAATCAAGG	2160 bp
	R: **TAGTGCGGCCGCTAG**CTAGTAGTCTGCAACCTCG	pDP1032-NheI
ple7	F: **CAGCAAGCTCCGAATTCGAGCTCGGTAC**ATGGAAGGCAAGGTGAGCTCC	958 bp
	R: *ATCCCCGGGTAC*CTAGATGACACTCCATGCGT	pDP1032-KpnI/SpeI
ple1	F: *CCGCACTAGT*ATGGCACCGTCTAATACAG	2124 bp
	R: **TTTCATCATCGGGTACGACTACCCGGGTCA**CTAAGCGCTGGTAGGCTTTG	pDP1032-KpnI/SpeI
linker (ple7-ple1)	F: TGTCATCTAG*GTACCCGGGGAT*CTGTAGTAG	
	R: ACGGTGCCAT*ACTAGTGCGG*CCGCTAGC	